MOLECULAR DYNAMICS PREDICTS THE SOLUTION CONFORMATIONS OF POLY-L-LYSINE IN SALT SOLUTIONS

by

Liqi Feng

B.S. in Chemistry, Nankai University, 2011

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

M.S. in Chemistry

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Liqi Feng

It was defended on

April 16, 2014

and approved by

Sanford A. Asher, Department of Chemistry

Jeffry Madura, Depatment of Biochemistry, Duquesne University

Lillian Chong, Department of Chemistry

Thesis Advisor: Sanford A. Asher, Department of Chemistry

MOLECULAR DYNAMICS PREDICTS THE SOLUTION CONFORMATIONS OF POLY-L-LYSINE IN SALT SOLUTIONS

Liqi Feng, M.S.

University of Pittsburgh, 2014

Ultraviolet resonance Raman (UVRR) studies recently discovered that increasing concentrations of $NaClO_4$ increase the fraction of α-helical conformations of poly-L-lysine (PLL) in water solutions. In contrast, this α-helical content increase does not occur for NaCl solutions. We used enhanced sampling molecular dynamics to explore the conformational space of PLL and to examine the effect of ions on PLL conformation. The free-energy landscapes of PLL in solutions were determined using the simulation data. The simulation results were also used to develop a molecular picture of ion-PLL interactions as well as the impact of ions on peptide hydration. The examination of pair interaction energies reveals the mechanisms whereby ions stabilize PLL conformations. ClO_4^- increases the α-helix conformation by decreasing the hydration of the peptide backbone which stabilizes the α-helical intramolecular hydrogen bonds (H-bonds). This occurs because of the relatively large ClO_4^- size and its tetrahedral structure. In contrast, the smaller Cl^- negligibly impacts the peptide backbone hydration and does not stabilize intramolecular H-bonds. In summary the results reported here support the experimental observations and provide a molecular picture of the role ions play in PLL conformations in aqueous salt solutions.

TABLE OF CONTENTS

PREFACE . . . viii

1.0 INTRODUCTION . . . 1

1.1 Stability of Protein Secondary Structures . . . 1

1.2 Molecular Dynamics and Metadynamics . . . 2

2.0 MOLECULAR DYNAMICS PREDICTS THE SOLUTION CONFORMATIONS OF POLY-L-LYSINE IN NACL AND $NACLO_4$ SOLUTIONS . . . 7

2.1 Introduction . . . 7

3.0 METHODS . . . 9

3.1 Metadynamics simulation . . . 9

3.2 Simulation detail . . . 9

3.3 Ψ-angle distribution . . . 10

3.4 Classical MD simulation . . . 11

3.5 Pair interaction energy distribution . . . 11

3.6 Radial distribution function . . . 12

4.0 RESULTS AND DISCUSSION . . . 14

4.1 Free energy surface . . . 14

4.2 Ψ-angle distribution . . . 15

4.3 Radial distribution . . . 17

4.4 Pair energy calculations . . . 18

5.0 CONCLUSION . . . 29

6.0 FUTURE WORK . . . 31

BIBLIOGRAPHY 33

LIST OF TABLES

1 Simulation parameters for history-dependent potentials 3
2 Summary of %α-helical-like Ψ angles 21

LIST OF FIGURES

1 Two-dimension free energy landscapes of n-pentane in vacuum. Results are shown as the energies along two dihedral angles (C1-C2-C3-C4 and C2-C3-C4-C5) . Lower energies are colored red and higher energies are colored yellow. Region in white are high in energy because of the steric clashes. From A to E, they are the results from simulations 1 to 5. 5

2 Free energy landscape of n-pentane in vacuum using DFT calculation by Jan Martin[14]. ϕ_1 and ϕ_2 are the dihedral angles along C1-C2-C3-C4 and C2-C3-C4-C5 respectively. 6

3 Ideal structures used to define the reference coordinates in determining the RMSD distance. A) ideal PPII, B) ideal α-helix and C) ideal helix-turn-helix. 13

4 Conformational free energy landscapes of PLL in solution. A) PLL in H_2O, B) PLL in 0.8 M NaCl and C) PLL in 0.8 M $NaClO_4$. Three axes are RMSD (PPII), RMSD (Helix) and RMSD (helix-turn-helix). Energy values of each free energy landscape are divided into 4 levels colored red, green, blue and white, respectively. 20

5 Ψ-angle distribution of PLL in solutions. A) PLL in H_2O, B) PLL in 0.8 M NaCl and C) PLL in 0.8 M $NaClO_4$ 21

6 Radial distribution functions, g(r) between A) PLL backbone atoms and water O and B) PLL backbone atoms and water H. g(r) for 1PCL, 2 PCL, 3PCL, 1NACL, 2NACL and 1WAT are colored red, blue, black, pink, yellow and green, respectively. 22

7 Radial distribution functions, A): g(r) of Cl of ClO_4^- with respect to the peptide backbone of 1PCL (red), 2PCL (blue) and 3PCL (black) and B): g(r) of Cl^- with respect to the peptide backbone of 1NACl (pink) and 2NACL (yellow). 22

8 Radial distribution function, g(r) between A) the peptide backbone of 2PCL and water O and B) the peptide backbone of 2PCL and water H. 2PCl in 0.8 M $NaClO_4$, 0.8 M NaCl and H_2O are colored blue, red and black, respectively. 23

9 Pair energy distribution (PED) between peptide structures and ions. A) PED between 2PCL and Cl^-, and PED between 2PCL and ClO_4^-, B) PED between 1PCL and Cl^-, and PED between 1PCL and ClO_4^- 24

10 Spatial distribution of anions around PLL peptide. Anions' positions (Cl^- or ClO_4^-) are shown by small red dots around the peptide. Cl^- with a PI energy less than or equal to -46 kcal/mol and ClO_4^- with a PI energy less than or equal to -22 kcal/mol are shown. 25

11 Detailed PEDs for extended (1PCL) and folded (2PCL) structures in salt solutions. A) PED of 1PCL in NaCl, B) PED of 1PCL in $NaClO_4$, C) PED of 2PCL in NaCl and D) PED of 2PCL in $NaClO_4$ 26

12 The PED for 2PCL in $NaClO_4$ from a population of anions that have a PI with backbone over 15 kcal/mol. 27

13 Representative peptide structures with anions. Peptide are drawn as Licorice and the ions are using ball and stick representation. Secondary structure is highlighted by ribbon. 28

PREFACE

I would like to express my gratitude to my thesis advisor, Professor Sanford Asher, for the knowledge, support and insight that he generously provided during my graduate study at the University of Pittsburgh. Dr. Asher encourages me to have an in-depth understanding of every detail in a scientific research project from which I benefitted.

I would also like to thank Professor Jeffry D. Madura. Dr.Madura guided me through the research and gave me a lot of helpful advice on how to train myself as a scientist. I have learned a lot during the stay in his lab. Without his suggestions and assistance, I would not be able to understand the thesis project.

A special thanks goes to one of my thesis committee, Professor Lilian Chong who has been patient and considerate in my graduation. She also inspired me in pursuit of a project with a computational focus.

I gratefully acknowledge all the members of Asher research group and all the members of Madura research group. I not only gain much insight on my project but also have a great time being with all my lab mates.

Finally, I sincerely thank my Mom, Jinxin Li, for understanding me and giving me unlimited support in my graduate study.

1.0 INTRODUCTION

1.1 STABILITY OF PROTEIN SECONDARY STRUCTURES

Proteins perform key functions within living organisms, including catalyzing biochemical reactions, transporting signals and constructing muscles. A protein's function depends heavily on its structure. The structure of a protein depends on the conformation of two angles, Φ and Ψ, as well as the surrounding environment. There are four levels used to describe a protein structure. The primary structure of a protein is the amino acid sequence from N terminus to C terminus. The secondary structure is the spatial arrangement of its backbone atoms. The association of secondary structural elements of the same chain of amino acid residues results in the tertiary structures. The assembly of several protein chains leads to the quaternary structure. Identifying factors in the stability of secondary structures is an important step in understanding protein structure. The secondary structures of a protein/peptide can be classified into the following categories based on the sequence of Ψ and Φ angles of peptide bonds: α-helix, π-helix, 3_{10}-helix, 2.5_1-helix, PPII and random coil. The main stabilizing force for α-helix, π-helix, 3_{10}-helix, 2.5_1-helix and PPII is the hydrogen bonding between backbone carbonyl oxygen and amide hydrogen groups.

Many different experimental techniques have been applied to probe the protein energy landscape and stability of protein structures including X-ray crystallography, infrared spectroscopy (IR), circular dichroism (CD), mass spectrometry and ultraviolet resonance Raman spectroscopy (UVRR). A recent research on poly-L-lysine (PLL) using UVRR thoroughly examines the specific ion effect on the stability of secondary structures[1,2]. Ma et al. obtain the fractions of different secondary structures of PLL by deconvolution of the $AmIII_3$ band in spectra. They discovered that addition of $NaClO_4$ increased the stability of α-helical like

structures of PLL in solution. However, the same concentration of NaCl did not display a similar effect. The dominant structure of PLL in concentrated NaCl solution is the same as the dominant structure of PLL in pure water solution. Bello et al. also observed increased helix content of PLL and methylated PLL with addition of $NaClO_4$[3]. Asciutto et al. have attributed increased helicity to ClO_4^-'s ability to dehydrate the peptide backbone by molecular simulation study on an alanine-based peptide AP[4]. The high affinity of large soft anions like SCN^- for the protein backbone is again observed in a molecule dynamics study by Petrava et al.[5] where they look at the densities of different anions around the peptide backbone. In those simulation studies, molecular dynamics results gave detailed information on ion-peptide interaction including a molecular picture of how ions interact with peptide functional groups, the density distribution of ions around the segment of peptide of interest and the affinity of ions to the segments of peptide.

1.2 MOLECULAR DYNAMICS AND METADYNAMICS

Molecular dynamics (MD) is the application of Newton's 2^{nd} law of motion to simulate the movement of particles. It has been used to understand the structure and function of biological macromolecules[6]. MD simulations provide detailed information on atom motions as a function of time which can be difficult to achieve by experiment. In brief, MD has been applied to simulate biological systems in order to sample the configuration space of biomolecules, to obtain a description of the biological system at equilibrium and to examine the actual dynamics of the biological system. To determine an accurate description of atomic motions as a function of time, one needs to use a time step on the order of femtoseconds (10^{-15} s) in classical MD simulation[7]. However, events of interest like unfolding of a small peptide often happen on a much longer timescale, $\sim$ miliseconds (10^{-3} s) to seconds, that may require an impractical amount of computational resources. To overcome the timescale problem, enhanced sampling techniques are developed to accelerate MD simulation[8,9]. Metadynamics is one of the enhanced sampling methods in which sampling is facilitated by history-dependent potentials[7]. The history-dependent potential acts on a selected number of degrees of free-

dom to discourage the system from revisiting configurations that have been sampled. Often, the history-dependent potential is a Gaussian function. At the end of the metadynamics simulation, one is able to construct the free energy surface of the simulated system as the function of selected degrees of freedom.

The efficiency and accuracy of metadynamics simulation depends on several parameters including the width of the history-dependent potential, the height of the history-dependent potential, the frequency of the history-dependent potential deposition and the selected degrees of freedom[10]. In order to find the right set of these parameters for metadynamics simulation, we investigated the impact of these parameters on the free energy landscape of pentane in vacuum. To obtain a well-represented free energy landscape, one needs to select the slow degrees of freedom that govern the conformational change. In the case of pentane, the dihedral angles along C1-C2-C3-C4 and C2-C3-C4-C5 correspond to the slowest degrees of freedom. To figure out the appropriate parameters for history-dependent potentials, we performed the following metadynamics simulations with CHARMM36 force field parameters for alkanes[11,12] in NAMD2.9[13]:

Table 1: Simulation parameters for history-dependent potentials

Simulation#	hillWeight (kcal/mol)	hillWidth (degree)	simulation time (ns)
1	0.3	6	14
2	0.3	40	0.33
3	0.3	40	0.36
4	0.6	40	0.016
5	0.6	40	0.02

where hillWeight stands for the height of the history-dependent potential and hillWidth means the width of the history-dependent potential.

Laio et al[10] proposed that the required simulation time follows:

$$t_{total} = \frac{\Delta G}{w} \tau_G (\frac{s}{\delta s})^d$$

where the terms in the expression are:

ΔG is the estimated energy barrier between two states. In the case of pentane, ΔG=6 kcal/mol.

w is the hillWeight of the history-dependent potential.

τ_G is the frequency of the history-dependent potential deposition. In the case of pentane, τ_G=200 fs.

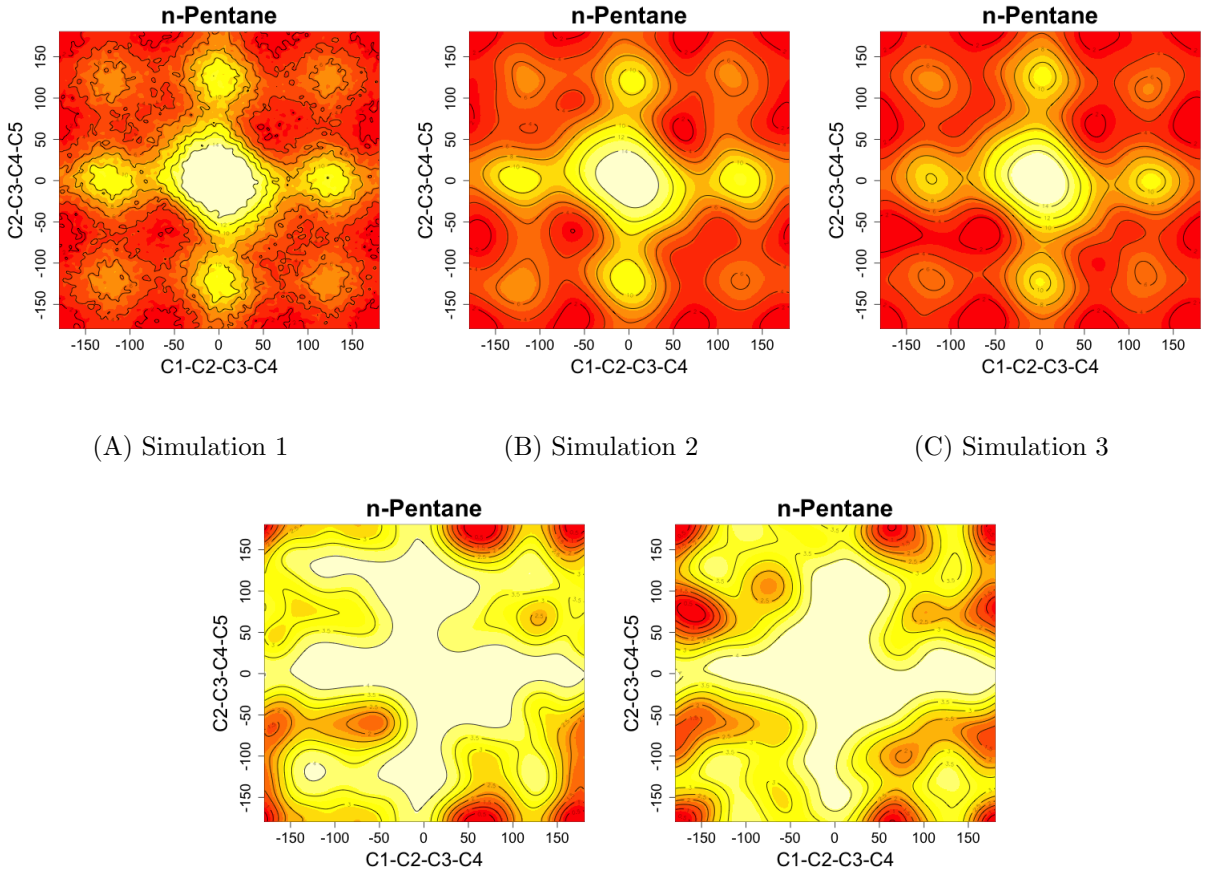
δs is the hillWidth of the history-dependent potential.

s is the the magnitude of the slow degree of freedom. In the case of pentane, s=360 degree.

d is the dimension of the slow degrees of freedom. In the case of pentane, d=2.

Liao et al. also calculated that the optimal choice for $\frac{s}{\delta s}$ in a two-dimensional search to be 0.1. This value was used in simulations 2 to 5. The simulation time for simulations 1, 2 and 4 are the same as the suggested values by Laio et al.. Simulations 2 and 4 were run longer to see how the length of a simulation affects the free energy landscape. These are labeled simulations 3 and 5 in Table 1. The results are shown in Figure 1.

Comparing Figure 1A to Figure 1B, we can conclude that using a smaller hillWidth of the history-dependent potential would not only slow down the simulation but also increase the ruggedness of the free energy landscape. With increased hillWeight, the simulation can be done in a faster fashion but we lose the details of free energy landscape (Figure 1D). Therefore the information we obtain from this simulation is not desirable. With a prolonged simulation time, the underlying true free energy landscape got overfilled by the history-dependent potentials seen in Figure 1C and Figure 1E. The resulting free energy surface is less informative as we lose the detail. To assess the accuracy, we compare the result with the calculated free energy landscape using DFT method[14] (Figure 2). The result from simulation 2 matches the DFT result best and thus we can conclude the parameters in simulation 2 is reasonable for a metadynamics simulation of n-pentane in vacuum.



(A) Simulation 1 (B) Simulation 2 (C) Simulation 3

(D) Simulation 4 (E) Simulation 5

Figure 1: Two-dimension free energy landscapes of n-pentane in vacuum. Results are shown as the energies along two dihedral angles (C1-C2-C3-C4 and C2-C3-C4-C5) . Lower energies are colored red and higher energies are colored yellow. Region in white are high in energy because of the steric clashes. From A to E, they are the results from simulations 1 to 5.

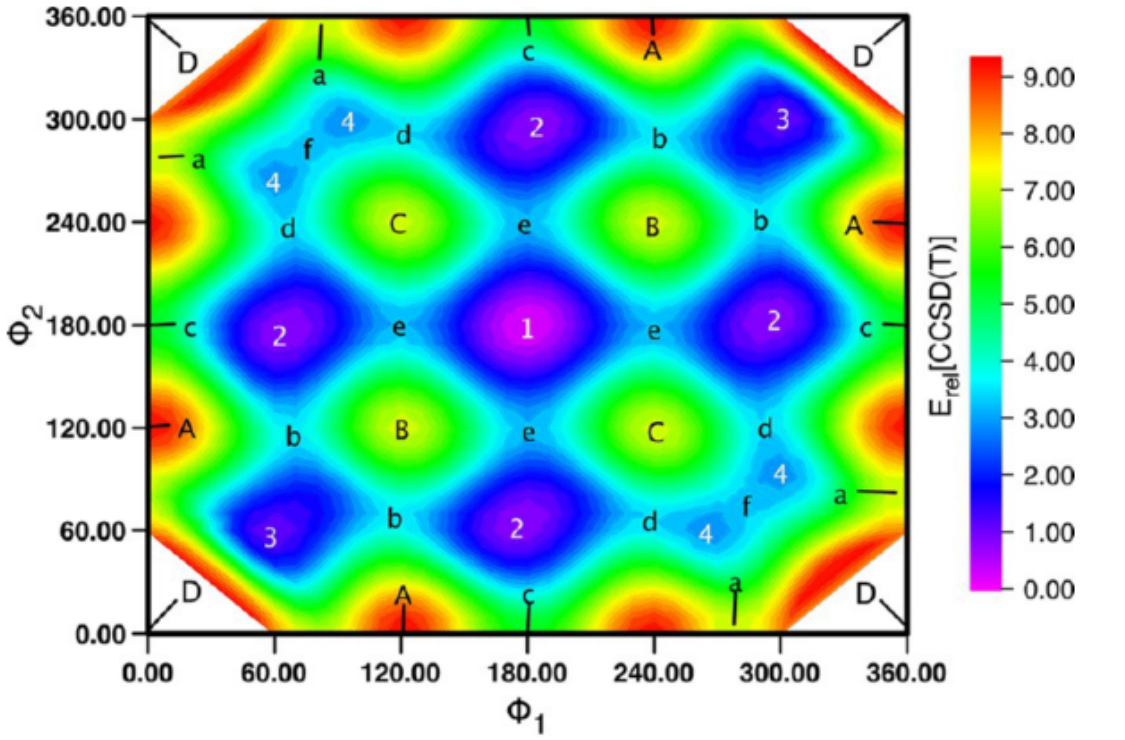


Figure 2: Free energy landscape of n-pentane in vacuum using DFT calculation by Jan Martin[14]. ϕ_1 and ϕ_2 are the dihedral angles along C1-C2-C3-C4 and C2-C3-C4-C5 respectively.

2.0 MOLECULAR DYNAMICS PREDICTS THE SOLUTION CONFORMATIONS OF POLY-L-LYSINE IN NACL AND $NACLO_4$ SOLUTIONS

2.1 INTRODUCTION

Most proteins spontaneously fold into their native structures. Thus, the protein primary sequence contains all the necessary information needed for folding. Unfortunately, there is still very little understanding of the mechanism(s) whereby proteins fold. This understanding of the mechanisms of folding would be significantly increased if the factors that stabilized protein conformations were defined. One of the factors defining protein structure involves ions in solution that form ion pairs and act as side chain counterions. These ions either stabilize or denature proteins.

The important role of ions has been known since at least 1888, from Frans Hofmeister's work that examined the stability of proteins in salt solutions[15]. In fact anions have been ordered in a list known as Hofmeister series in the order by which they destabilize the solution protein structure which generally involves unfolding. The series is: $SCN^- > ClO_4^- > I^- > ClO_3^- > Br^- > NO_3^- > Cl^- > CH_3CO_2^- > HPO_4^{2-} > SO_2^{2-}$. Until recently, it was hypothesized that these ions acted through their impact on water structure[16]. The current hypothesis is that the impact of these ions on protein solution structure is the result of the variation of the ions in their water affinities[17]. An ion pair with similar water affinities is most stable. Thus, ions with similar water affinities as charged protein side chain form the most stable ion pairs. In addition to water affinity, Collins et al. also pointed out that ion-protein binding is strengthened by a chelation involving multiple attachments[17].

There have been numerous computational studies on the role ions play in peptide con-

formation[5,18–21]. They focus on the binding affinities of different ions to the segments of peptides in various model systems. Of particular interest in this study are the results from Asciutto et al.[4]. They showed that ClO_4^- stabilized α-helical peptides by preferentially binding to the peptide backbone.

Recent UV resonance Raman (UVRR) studies of Ma. et al.[1,2], discovered that ClO_4^- dramatically stabilizes α-helical like solution structures of PLL. In a study of a long PLL peptide in $NaClO_4$ solution with UVRR spectroscopy, Ma et al. showed that the free energies of the α-helical-like solution structures are lower than the free energies of the extended solution structures[1,22].

In our work here, we examined the impact of ClO_4^- and Cl^- on the solution conformation of a short poly-L-lysine (PLL) peptide. We investigated the mechanism of α-helical stabilization of PLL by the analysis of the results of enhanced molecular dynamics simulations (eMD). We determined the conformational space of PLL in salt solutions and developed an understanding of the role of ClO_4^- on peptide conformation. We found that ClO_4^- can simultaneously bind to the side chains and the peptide backbone and can protect intramolecular hydrogen bonds (H-bonds) from hydration. This stabilizes α-helix-like structures. In contrast, Cl^- has little impact on the hydration of the peptide backbone. Thus Cl^- does not stabilize α-helical-like structures.

3.0 METHODS

3.1 METADYNAMICS SIMULATION

To efficiently sample the conformational space of PLL in solution, we used the enhanced sampling molecular dynamics technique, metadynamics. The metadynamics algorithm, as explained by Henin et al.[23], samples the conformational space using history-dependent biasing function. Briefly, the metadynamics algorithm uses a history-dependent potential energy function to prevent the system from resampling the same conformation[24]. Typically, the history-dependent potential energy function is Gaussian. At the end of metadynamics simulation, the free energy landscape is constructed based on history-dependent potential energy functions that have been applied during the simulation[25]. The conformational search was carried out in a three-dimensional space defined by the RMSD distance of the peptide to an ideal α-helix, an ideal PPII helix and an ideal helix-turn-helix. Metadynamics simulations were performed utilizing the collective variable (COLVAR) module available in the NAMD 2.9 package[13].

3.2 SIMULATION DETAIL

The intra- and intermolecular potentials were defined by using the CHARMM27 force field with CMAP corrections[26]. Force field parameters for ClO_4^- were given by Baaden et al.[27] with atomic charges developed by Asciutto et al.[4]. Force field parameters for Na^+ and Cl^- were adapted from work by Joung et al.[28]. The TIP3P water model was used in all simulations[29]. For all simulations, the short-range van der Waals and electrostatic interactions

were truncated smoothly at a cutoff distance of 12 Å with a switching function applied over distances greater than 10 Å. Long-range electrostatic potentials were computed by the particle-mesh Ewald method[30]. Bonds involving H atoms were fixed using the RATTLE algorithm[31,32]. NPT simulations were carried out with a time step of 2 fs. The pressure was maintained at 1 atm using a Langevin piston algorithm[33] and the temperature was maintained at 298 K using a damped Langevin dynamics. The atomic positions during the simulation were saved every 1000 steps.

The collective variables used in the metadynamics simulation of PLL were defined as the distance root-mean squared deviation (RMSD) of the backbone atoms of the peptide with respect to an ideal PPII, an ideal α-helix and an ideal helix-turn-helix (HTH), respectively. These ideal PLL structures, displayed in Figure 3, were built using the Ramachandran Ψ and Φ angles of the classical PPII structure ((Ψ , Φ)= (140°, -65°)) and the classical α-helix structure ((Ψ , Φ)= (-39°, -65°)). For the ideal HTH structure, the backbone Ψ, Φ angles for the first and last seven residues were set to (-47°, -57°) and the backbone Ψ, Φ angles for the middle three residues were set to (30°, -60°). Each collective variable was discretized over a range of 0 Å to 14 Å with a bin size of 0.1 Å. External Gaussian potentials with a height of 2 kcal/mol were applied every 100 steps.

We simulated a 17-residue long PLL peptide in its zwitterionic form (0.8 mg/mL) immersed in a water box with dimension 80×80×80 $Å^3$ in 0.8 M NaCl or 0.8 M $NaClO_4$. The salt system studied contained one peptide, 15700 waters, 230 Na^+ and 247 Cl^- or 247 $ClO_4{}^-$. The pure water system contained one peptide, 16160 waters and additional 17 Cl^- to ensure neutralized system. Simulations were run for 670 ns for PLL in 0.8 M $NaClO_4$, 640 ns for PLL in 0.8 M NaCl and 640 ns for PLL in pure water.

3.3 Ψ-ANGLE DISTRIBUTION

The Ramachandran Ψ-angle distribution was generated by calculating a histogram of the observed frequency of Ψ angles of the structures found in the metadynamics simulation. The resulting biased probabilities were adjusted by multiplying them by $e^{-\beta\Delta G}$ where ΔG was

obtained from the free energy landscape. Ψ angles of the peptide were measured every 8000 steps.

3.4 CLASSICAL MD SIMULATION

To examine the PLL hydration and ion-peptide interactions, we performed 50 ns of fixed-backbone classical molecular dynamics (cMD) simulations using NAMD 2.9 package. These simulations used the same force field and parameters as used in the metadynamics calculations. The atomic positions were collected every 1000 steps over the 50 ns simulation.

3.5 PAIR INTERACTION ENERGY DISTRIBUTION

Pair interaction (PI) energy calculations were performed utilizing the CHARMM[34,35] 35b3 package using the same force field and parameter as used in the metadynamics simulations. The PI energy between two groups A and B is defined as[36]:

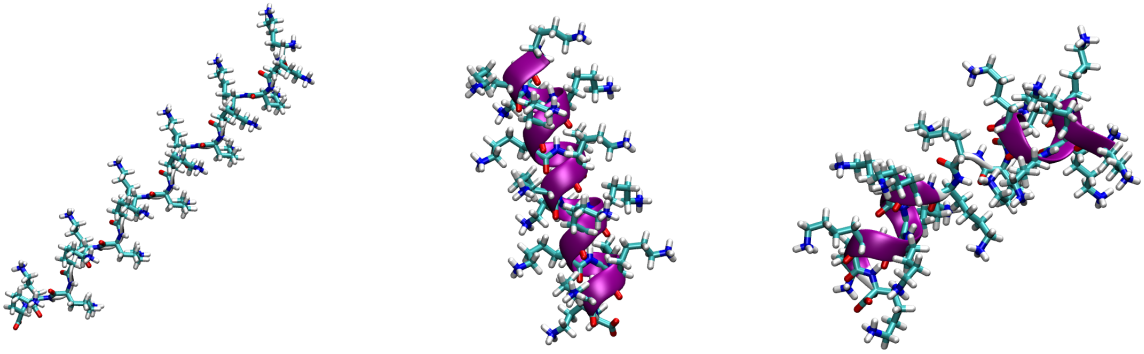
$$E(A,B) = E(AB) - [E(A) + E(B)]$$

where E(AB) represents potential energy of the complex AB, E(A) and E(B) are the potential energies of isolated A and B, respectively. In our analysis, A is the ion (Cl^- or $ClO_4{}^-$). B is either the peptide, the peptide side chain or the peptide backbone. We focus on PI energy between a segment of the peptide and the ion as well as between the peptide and the ion.

PI calculations of ions situated within 12 Å from the peptide utilized snapshots taken every 50000 steps in the fixed-backbone cMD simulations. Normalized PI distributions were determined by creating a histogram of PI energies.

3.6 RADIAL DISTRIBUTION FUNCTION

The radial distribution function, g(r) is the probability of finding a particle at a specific distance from a reference particle. We use g(r) to characterize the coordination of water and the coordination of anions around the peptide backbone. g(r) was computed using PTRAJ[37] available in AmberTools. The g(r) was calculated from snapshots taken every 10000 steps in the fixed-backbone cMD simulations.



(A) ideal PPII (B) ideal α-helix (C) ideal helix-turn-helix

Figure 3: Ideal structures used to define the reference coordinates in determining the RMSD distance. A) ideal PPII, B) ideal α-helix and C) ideal helix-turn-helix.

4.0 RESULTS AND DISCUSSION

4.1 FREE ENERGY SURFACE

Figure 4 shows the calculated conformational free energy landscapes of PLL in water and in 0.8 M $NaClO_4$ and NaCl plotted using the R package[38]. The free-energy landscapes spans a reduced-dimension space that enables us to make direct comparisons of the relative free energies of different structures. The ideal structures used to establish the three-dimensional space are shown by the colored spheres in Figure 4, where the red sphere represents the ideal PPII conformation, the blue pentagon represents the ideal α-helix conformation and the green square represents the ideal helix-turn-helix conformation.

The free energy values are divided into 4 energy levels, colored as red, green, blue and white ordered in increasing energy. The location of structures at relatively low free energies are indicated by the small black spheres. Representative structures at these energies were extracted from simulation trajectories using the coordinates of black spheres in the free energy landscape. Representative structures with these low free energies are shown next to the black spheres.

Figure 4A, showing the free energy landscape of PLL in water, has a broad energy well with conformations that show a small RMSD ($\sim$ 2-8 Å) from the PPII conformation, but a large RMSD ($\sim$ 4-10 Å) compared to the α-helix confirmation. The edge of the well is mainly shown in green. Within this energy well, the extended structure (1WAT) has the lowest free energy.

Figure 4B shows that the energy landscape of PLL in 0.8 M NaCl has two low free energy wells colored green. These low free energy wells are associated with an extended structure (1NACL) and a coil structure (2NACL). The free energy landscapes seen in Figure 4A and

4B indicate α-helical structures are not favorable in pure H_2O or NaCl solution.

Figure 4C shows that the free energy landscape of PLL in $NaClO_4$ differs from that in water and in NaCl . PLL in $NaClO_4$ shows three energy wells for conformations that are similar to α-helix conformations with a small RMSD from the α-helix structure. Two representative structures from these low free energy wells are a helical structure with a turn segment (2PCL) and a helical structures with an extended segment (3PCL). A low free energy well for extended conformations is observed in which an extended structure (1PCL) is selected to represent conformations in this well.

Thus, ClO_4^- lowers the free energy of α-helical conformations in agreement with experimental UVRR results[22]. Their results demonstrated that the free energies of α-helical-like structures were lower than the free energies of extended structures such as the PPII conformations. They also showed that NaCl does not stabilize α-helical structures[2] which is in agreement with our simulation result.

4.2 Ψ-ANGLE DISTRIBUTION

Figure 5 shows the calculated Ramachandran Ψ angle distributions of PLL in solutions. The Ψ-angle distribution indicates the relative populations of the PLL solution secondary structures. Ψ angles in the α-helical structure region ($\sim$ -60° to $\sim$ -25°) are colored red, while Ψ angles in the extended structure region ($\sim$ -180° to $\sim$ -160° or $\sim$ 110° to $\sim$ 180°) are colored blue. Angles associated with turn structures and other random coil structures are in black.

Extended structures dominate the conformations of PLL in NaCl with very small fraction of α-helical-like structures (Figure 5B). The α-helical fraction is comparable to the extended fractions in pure water (Figure 5A). In contrast, Figure 5C shows that PLL in $NaClO_4$ has an increased α-helical fraction. Taking into π-helix and the broader α-helical peak in distribution, the total fraction of α-helical-like structures is 36%. Ma et al.[1] showed that at pH 3 and at 20°C the fraction of α-helical like conformations was 60% for PLL in 0.8 M $NaClO_4$ (see Figure 3 of reference 2).

We define that a Ψ angle within a helical structure is a true helical Ψ angle. A helical structure has to have more than 3 successive helical Ψ angles. As Table 2 shows, most of the helical Ψ angles of PLL in **pure water** are not in helical structures. These random helical Ψ angles correspond to random coil structures. In contrast, 18.8% out of 27.4% of helical Ψ angles of PLL in $NaClO_4$ are in helical structures. Comparing the fraction of true helical angles among three solutions, the addition of ClO_4^- does show an increased fraction of α-helical-like structures.

Krimm et al. have shown that PPII structures are the most stable structure for PLL in pure water using circular dichroism (CD) spectroscopy[39]. This result is later confirmed by Rucker et al. by measuring the CD spectra of the 7-residue long PLL[40]. The discrepancy between our work and their CD spectroscopy result may due to the use of force field parameters. It is noticed that the backbone potential parameters in CHARMM27 force field would bias towards α helical structures[41]. Therefore Best et al. optimized the force field parameters using the latest experimental data which led to the CHARMM36 force field. The CHARMM36 force field includes terms that correct the preference for α-helical structures. Best et al. applied the CHARMM36 force field parameters in the case of an alanine-based peptide and the simulation yielded a decreased fraction of α-helical structures compared to the result from a simulation using CHARMM27 force field parameter. The CHARMM36 simulation result matched the experimental result by NMR. We will test this hypothesis by carrying out a new set of metadynamics simulations using CHARMM36 force field parameters.

The increased true α-helical fraction in $NaClO_4$ solution does not result from anion screening since NaCl should similarly screen. To understand the increased PLL α-helical fraction in $NaClO_4$ solution, we investigated the hydration of the PLL backbone as well as anion-peptide interactions by fixed backbone-atoms cMD of the most stable conformations found in the metadynamics simulation trajectories. As shown in Figure 4, six low energy structures are found that we labeled 1WAT, 1NACL, 2NACL, 1PCL, 2PCL and 3PCL, respectively. 1WAT, 1NACL and 1PCL are extended structures. 2PCL and 3PCL are defined as folded structures in our discussion. To study the stability of α-helical structures in pure water and 0.8M NaCl we also simulated 1PCL and 2PCL in 0.8 M NaCl and in pure

water.

4.3 RADIAL DISTRIBUTION

The radial distribution function, g(r) is the probability of finding an atom at a certain distance from a reference location. We calculated g(r) for the water O and H atoms to the peptide backbone. As shown in Figure 6, this g(r) indicates the hydration around the peptide backbone. The first peak in the g(r) for the water O occurs at $\sim$ 3 Å for all representative structures (see Figure 6A). Similarly, the first peak in the g(r) for water H occurs at $\sim$ 2 Å for all representative structures (see Figure 6B). The first hydration shell of the peptide backbone remains upon adding salts. A significant decrease in the value of g(r) occurs for the 2PCL and 3PCL partially α-helical structures at $\sim$ 5 Å. This indicates a less hydrated peptide backbone for folded structures in $NaClO_4$. This less hydrated peptide backbone for folded structure 2PCL does not occurs for simulations of 2PCL in 0.8 M NaCl and pure H_2O. Detailed results are presented in following discussion. Among extended structures (1PCL, 1NACl and 1WAT), a slight decrease in the value of g(r) for 1PCL also occurs at $\sim$ 5 Å. There is no significant decrease observed of g(r) values $\sim$ 5 Å for representative structures in NaCl. Thus, Cl^- does not change the peptide backbone hydration.

This result is supported by plots in Figure 7A shows g(r) between peptide backbone and the Cl in $ClO_4{}^-$ in solution. The g(r) values at $\sim$ 5 Å in the case of PLL in $NaClO_4$ are greater than in NaCl indicates that $ClO_4{}^-$ is situated closer to the peptide backbone. Figure 7B shows, g(r) between the peptide backbone and Cl^-. The large g(r) values at $\sim$ 10 Å indicates that Cl^- interacts closely with peptide side chains because the 10 Å distance to peptide backbone corresponds to the length of an extended lysine side chain.

We compared the hydration of the 2PCL backbone, a partially α-helix, between 0.8 M $NaClO_4$ and NaCl, and pure water (Figure 8). $ClO_4{}^-$ clearly dehydrates the peptide as evident in the decreased g(r), specifically at 5 Å. g(r) is identical in the case of 0.8 M NaCl and pure H_2O. The higher probability of finding H_2O close to the peptide backbone in NaCl and pure water may destabilize intramolecular H-bonding which would destabilize

the α-helical structures because α-helical structure relies on intramolecular H-bonds.

4.4 PAIR ENERGY CALCULATIONS

To further investigate the basis for increased helicity of PLL in ClO_4^- over Cl^- the pair interaction energy distribution (PED) for 1PCL and 2PCL in 0.8 MNaCl and in 0.8 M $NaClO_4$ are calculated. Figure 9 shows that the PED for Cl^- and ClO_4^- significantly differ. The PED of 2PCL in NaCl (blue line in Figure 9A) has a peak (PDK1) at -62 kcal/mol with a much smaller peak below -100 kcal/mol indicating a small fraction of Cl^- binding very strongly for 2PCL. The PED of 1PCL in NaCl (blue line in Figure 9B) has a peak (PDK1) at -60 kcal/mol. There are 7-9 Cl^- ions with a PI energy less than -46 kcal/mol per peptide for both 1PCL and 2PCL in NaCl.

The PED for 1PCL and 2PCL in ClO_4^- (red line in Figure 9A and 9B) exhibits a broad peak (PDK2) at -41 kcal/mol. There are 16-17 ClO_4^- ions with a PI energy less than -22 kcal/mol per peptide for 1PCL and 2PCL in $NaClO_4$. The order of these values are in alignment with the results of Asciutto et al. which showed ClO_4^- prefers α-helical structures while Cl^- does not[4].

The PED results highlight the range of peptide-ion interaction (PDK1 and PDK2); however, these results do not give insight into where the ions are binding. To examine the spatial distribution of the ions about the peptide we plot the ion position with PI energy less than -46 kcal/mol for Cl^- and the ion position with PI energy less than -22 kcal/mol for ClO_4^- (see Figure 10). The spatial distribution of 2PCL in ClO_4^- (Figure 10A) shows dense clusters of ions near the lysine side chain and peptide backbone. This is in contrast to 2PCL in Cl^- (Figure 10B) where the dense clusters of ions are near the lysine side chains. The spatial distribution result agrees with the g(r) of Cl around the peptide backbone (Figure 7) where the maximum g(r) value occurs at 5 Å for PLL in $NaClO_4$ while the maximum g(r) value occurs at 10 Å for PLL in NaCl ($g(r)_{Cl-backbone}$ of 1PCL and 2PCL in 0.8 M NaCl are shown in supplementary material).

The ion-backbone (yellow) and ion-side chain (blue) contributions to the ion-peptide

(red) PED for an extended (1PCL) and a folded (2PCL) PLL in 0.8M NaCl and $NaClO_4$ are shown in Figure 11. As shown in Figure 11A and Figure 11C, the ion-side chain PED and the ion-peptide PED of the extended and folded structures in NaCl share similar features: 1) positions of stabilizing interaction peaks (peaks at $\sim$ 68 kcal/mol) are almost identical, 2) fraction at each pair energy is to a similar magnitude. In other words, ion-side chain PED almost overlaps with ion-peptide PED. This means ion-side chain interaction accounts for most of peptide-ion interaction for PLL in NaCl.

In contrast, in Figure 11B and 11D the ion-side chain and ion-peptide PED do not overlap. The ion-side chain PED shows a larger fraction of negative PI than the ion-peptide PED. This larger fraction of negative PI is offset by the increased fraction of positive ion-backbone PI. This is an indicative of ClO_4^- simultaneously interacting with the peptide backbone and with the peptide side chain. Figure 12 shows the PEDs of ClO_4 that were having strong repulsive interactions (positive PI) with the backbone of 2PCL. ClO_4^- interacting with the peptide backbone with a repulsive interaction energy over 15 kcal/mol are interacting with the peptide side chain (red line in Figure 12) with PI energies peak at -125 kcal/mol. The PED between those ClO_4^- ions and peptide side chains accounts for the increased fraction of negative PI energies seen in Figure 11D. The resulting PED of PLL peptide and ClO_4^- (blue line in Figure 12) gives rise to a broader PDK2 in Figure 9A.

What we conclude from the results in Figure 11 is that ClO_4^- simultaneously interacts with the side chain and peptide backbone while the Cl^- mainly interacts with the side chain. We illustrate this results in Figure 13. Figure 13A and 13B show ClO_4^- interacts with the lysine side chain and the PLL backbone at the same time. The ClO_4^- in Figure 13A pointed by an red arrow was interacting with two lysine side chains and the backbone N atom. The similar interaction motif is seen in Figure 13B where a ClO_4^- near the α-helical segment was in a position having interaction with both side chains and the peptide backbone. This dehydrated the peptide bone around the α-helical segment and protected the α-helical structure from the attack of waters. This is in contrast to Cl^- which is interacting with the lysine side chain as shown in Figure 13C and 13D. These results are consistent with the spatial distribution shown in Figure 10.

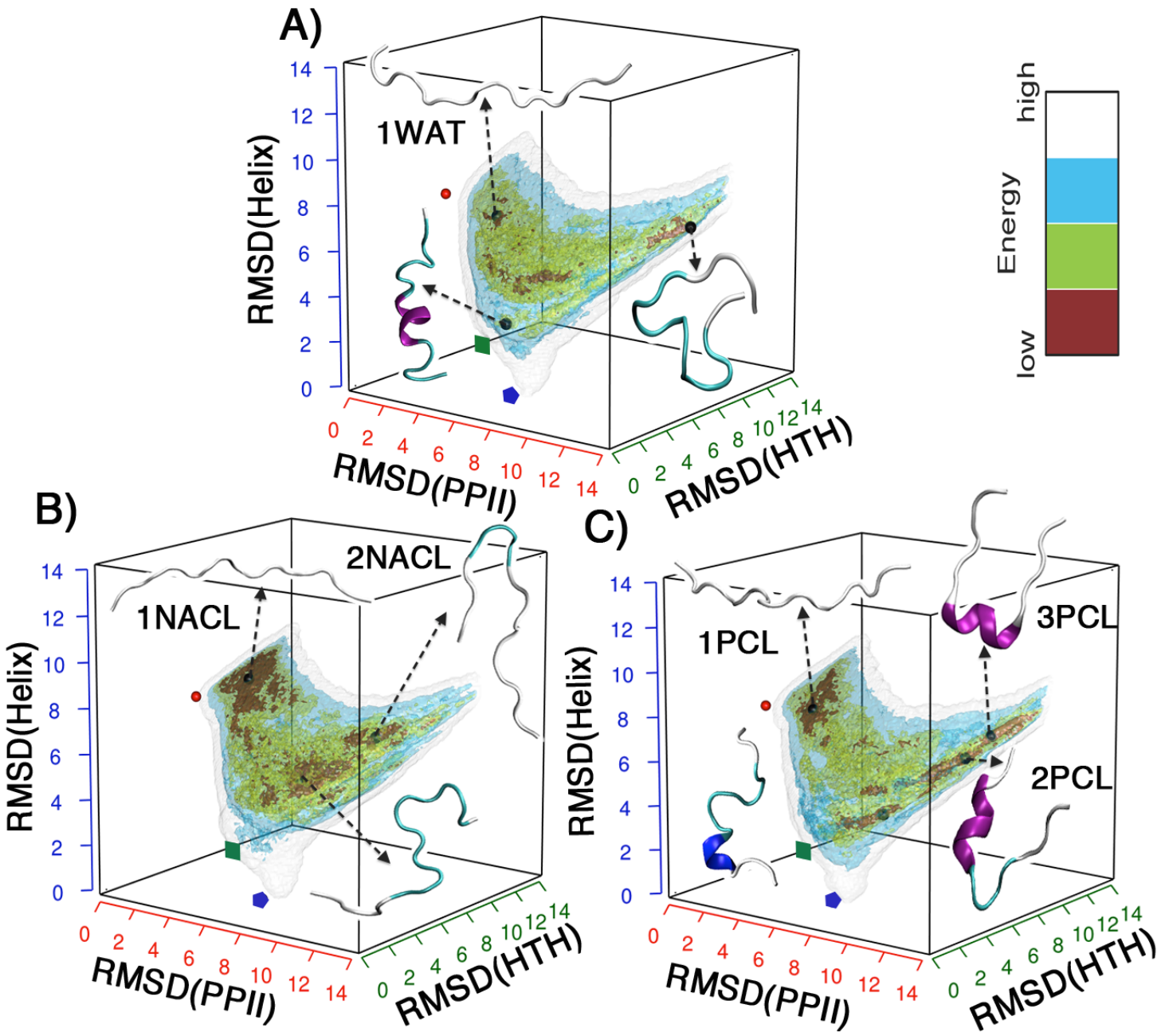


Figure 4: Conformational free energy landscapes of PLL in solution. A) PLL in H_2O, B) PLL in 0.8 M NaCl and C) PLL in 0.8 M $NaClO_4$. Three axes are RMSD (PPII), RMSD (Helix) and RMSD (helix-turn-helix). Energy values of each free energy landscape are divided into 4 levels colored red, green, blue and white, respectively.

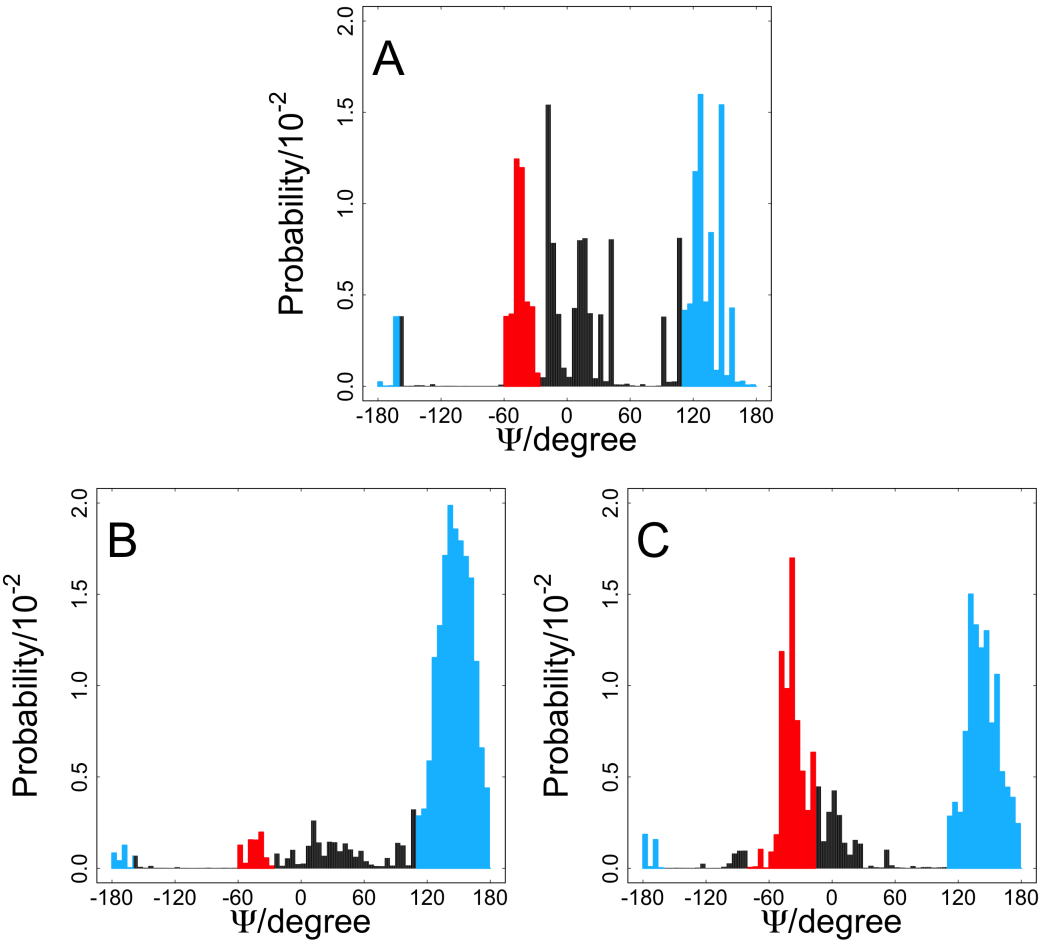


Figure 5: Ψ-angle distribution of PLL in solutions. A) PLL in H_2O, B) PLL in 0.8 M NaCl and C) PLL in 0.8 M $NaClO_4$

Table 2: Summary of %α-helical-like Ψ angles

Solution	%α-helical-like	%random (count <3)	%true helix (count ≥3)
H_2O	20.9	20.9	0.06
NaCl	3.6	3.6	0.08
$NaClO_4$	27.4	8.6	18.8

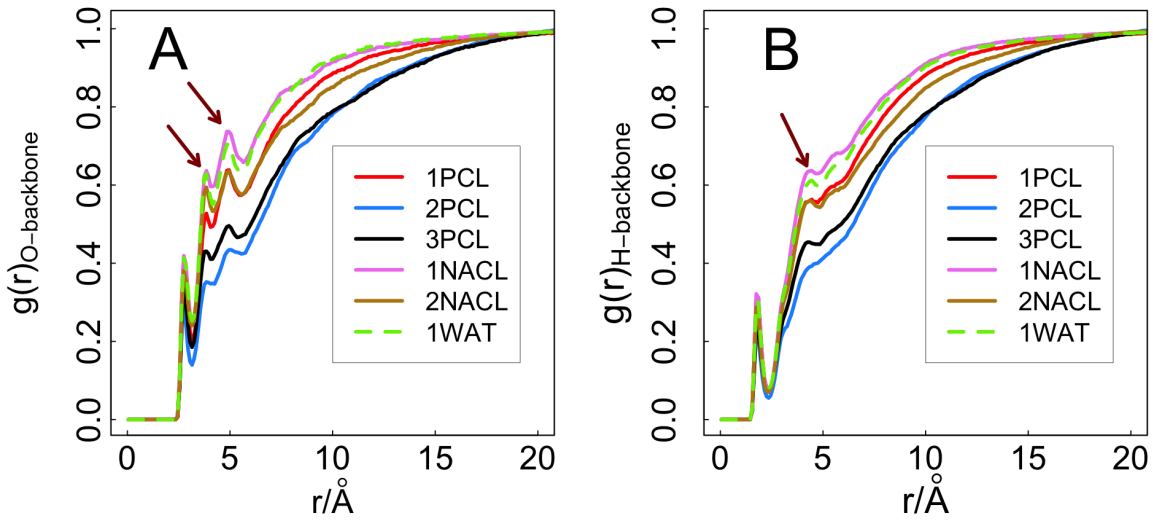


Figure 6: Radial distribution functions, g(r) between A) PLL backbone atoms and water O and B) PLL backbone atoms and water H. g(r) for 1PCL, 2 PCL, 3PCL, 1NACL, 2NACL and 1WAT are colored red, blue, black, pink, yellow and green, respectively.

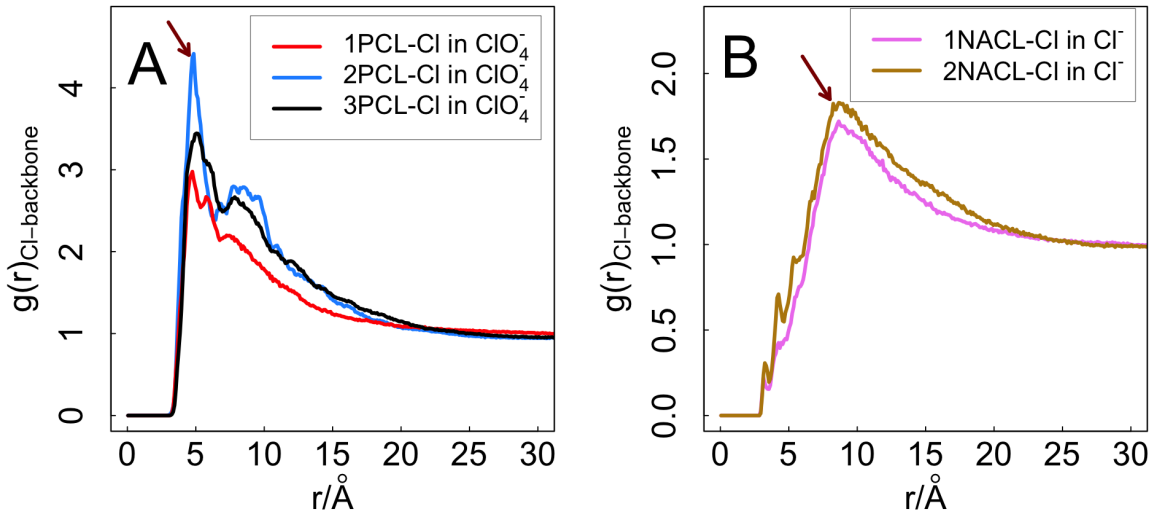


Figure 7: Radial distribution functions, A): g(r) of Cl of ClO_4^- with respect to the peptide backbone of 1PCL (red), 2PCL (blue) and 3PCL (black) and B): g(r) of Cl^- with respect to the peptide backbone of 1NACl (pink) and 2NACL (yellow).

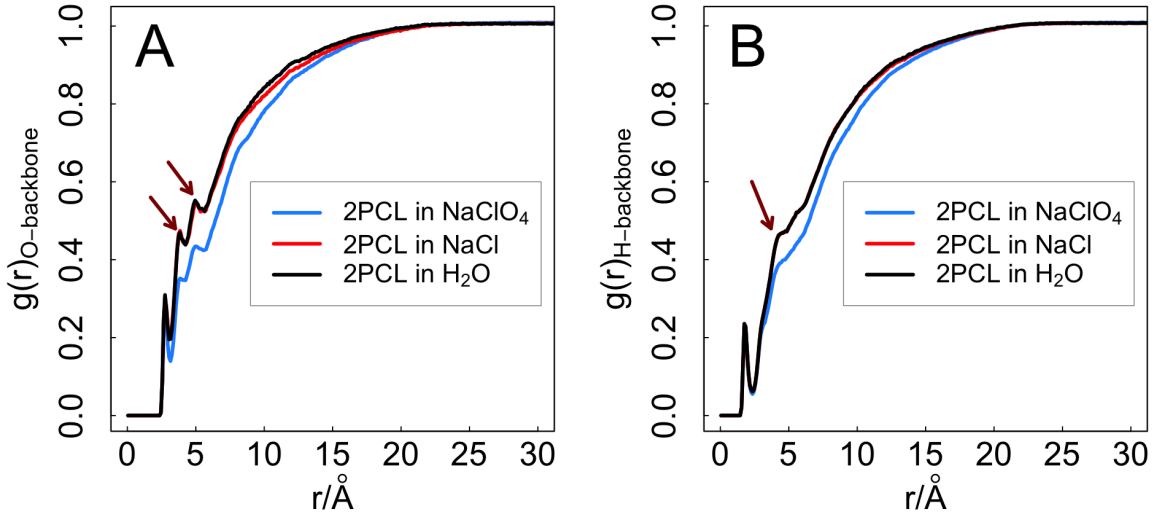


Figure 8: Radial distribution function, g(r) between A) the peptide backbone of 2PCL and water O and B) the peptide backbone of 2PCL and water H. 2PCl in 0.8 M $NaClO_4$, 0.8 M NaCl and H_2O are colored blue, red and black, respectively.

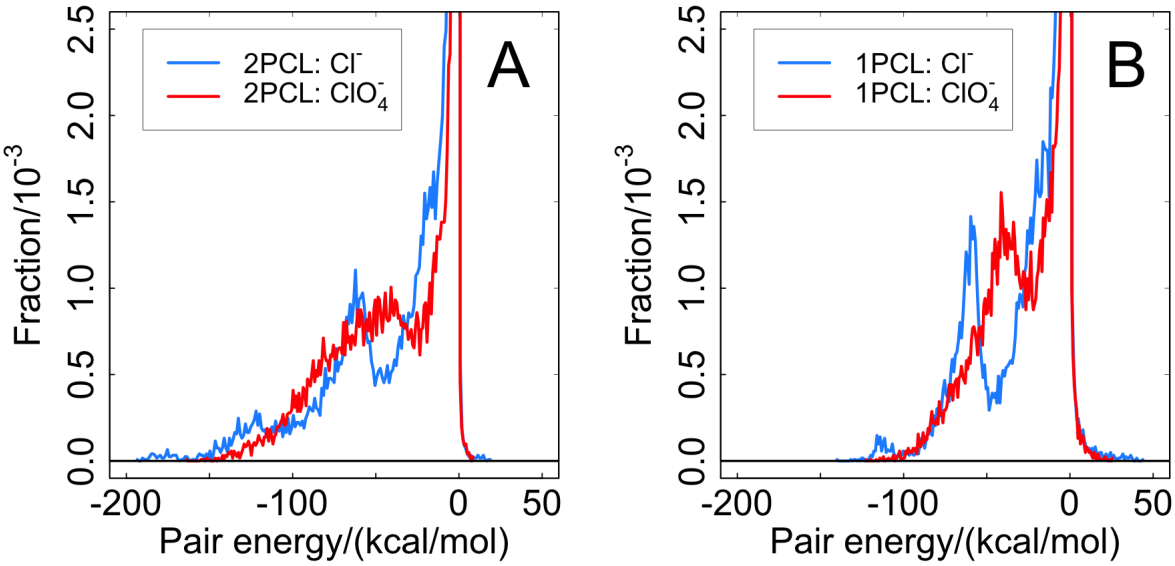
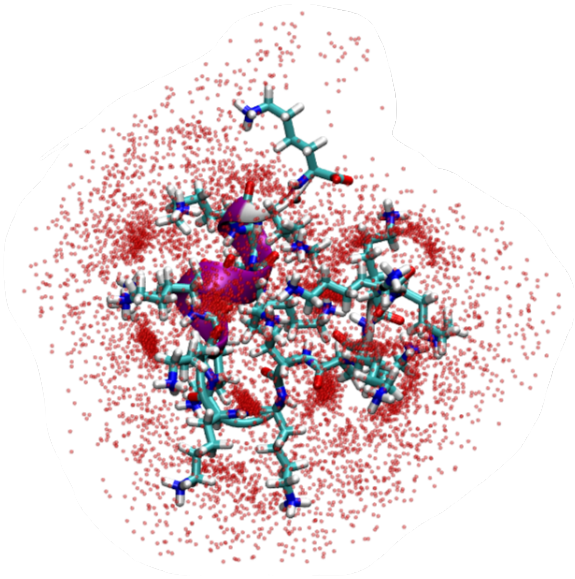
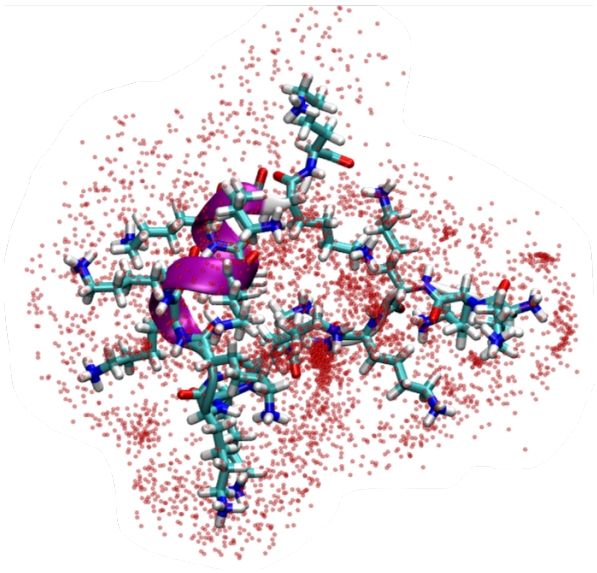


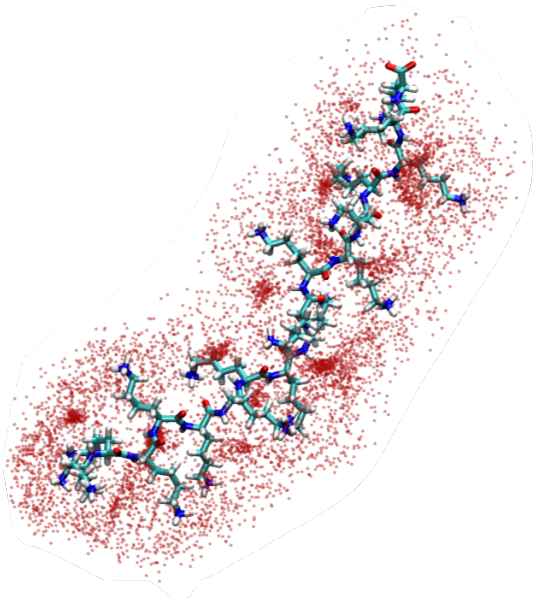
Figure 9: Pair energy distribution (PED) between peptide structures and ions. A) PED between 2PCL and Cl^-, and PED between 2PCL and $ClO_4{}^-$, B) PED between 1PCL and Cl^-, and PED between 1PCL and $ClO_4{}^-$



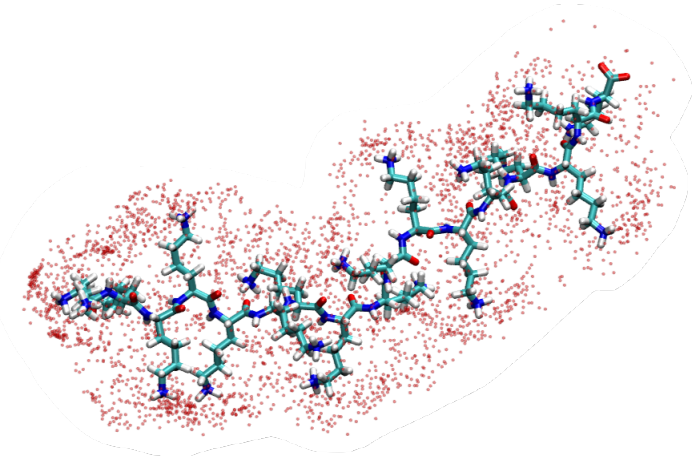
(A) ClO_4^- around 2PCL



(B) Cl^- around 2PCL



(C) ClO_4^- around 1PCL



(D) Cl^- around 1PCL

Figure 10: Spatial distribution of anions around PLL peptide. Anions' positions (Cl^- or ClO_4^-) are shown by small red dots around the peptide. Cl^- with a PI energy less than or equal to -46 kcal/mol and ClO_4^- with a PI energy less than or equal to -22 kcal/mol are shown.

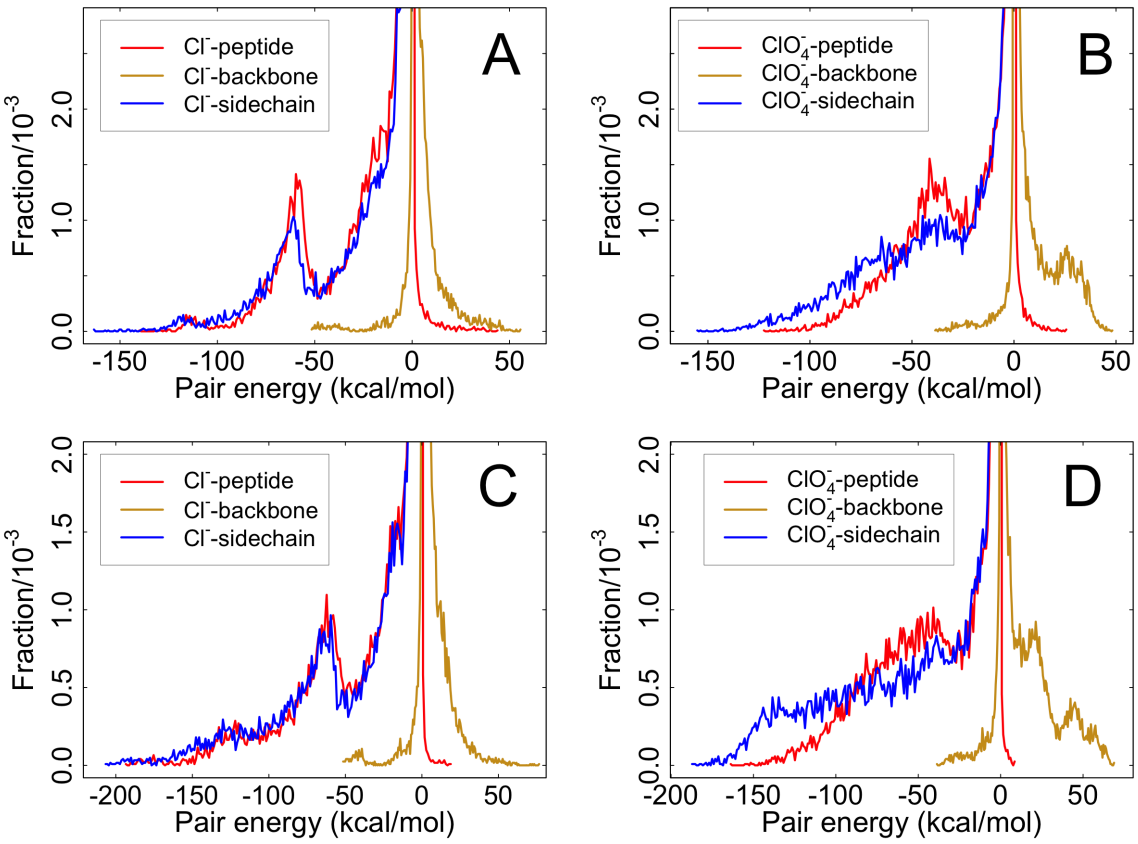


Figure 11: Detailed PEDs for extended (1PCL) and folded (2PCL) structures in salt solutions. A) PED of 1PCL in NaCl, B) PED of 1PCL in $NaClO_4$, C) PED of 2PCL in NaCl and D) PED of 2PCL in $NaClO_4$

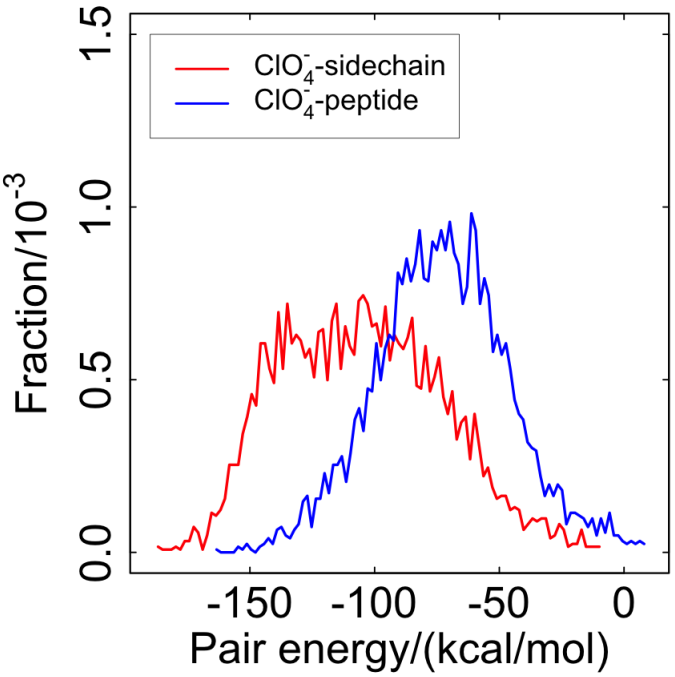
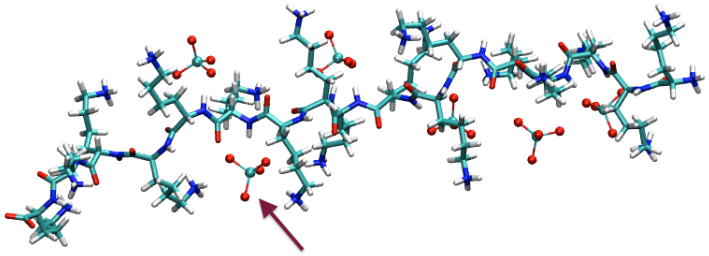
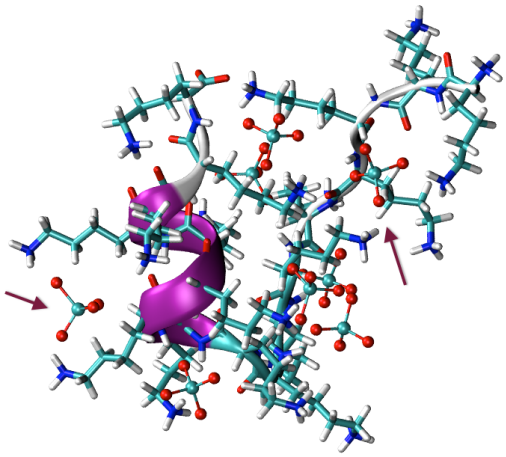


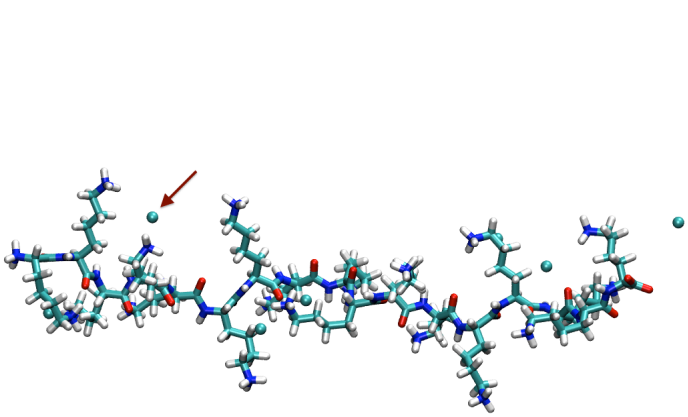
Figure 12: The PED for 2PCL in $NaClO_4$ from a population of anions that have a PI with backbone over 15 kcal/mol.



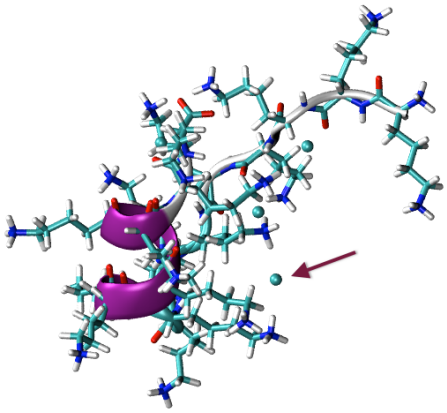
(A) 1PCL



(B) 2PCL



(C) 1NACL



(D) 2PCL in NaCl

Figure 13: Representative peptide structures with anions. Peptide are drawn as Licorice and the ions are using ball and stick representation. Secondary structure is highlighted by ribbon.

5.0 CONCLUSION

Preliminary metadynamics simulations have been performed to probe the conformational spaces of PLL in 0.8 M NaCl, 0.8 M $NaClO_4$ and pure water. Free energy landscapes were obtained and support the experimental results on specific-ion effect by Ma et al.[1,2,22] and Collins et al.[17] that is the affinity of ClO_4^- to PLL dehydrates the peptide backbone thereby stabilizing α-helical conformations. The same concentration of Cl^- does not exhibit a similar effect on PLL's conformation. This means salt screening effect can not account for this phenomenon. Radial distribution functions (RDF) and pair interaction (PI) energy calculations were determined to further investigate the stability of the α-helical structures in salt solutions. The RDF and PED results for ClO_4^- point out that ClO_4^- has significant ion-backbone and ion-side chain interactions. The RDF and PED results for Cl^- point out that the majority of Cl^- have only ion-side chain interaction. These results support our hypothesis that ClO_4^- decreases the hydration around the peptide and preserves intramolecular H-bonding leading to increased helix stability. Cl^- on the other hand does not alter the peptide hydration compared to bulk water and therefore yielding more extended and turn structures. Our Ramachandran Ψ-angle distribution qualitatively matches the finding by Ma et al. that an increased fraction of α-helical like structures is observed in high concentration of $NaClO_4$[2]. Asciutto et al. attributed the stabilization effect of $NaClO_4$ on α-helical structure to ClO_4^- interacting with the peptide backbone locally and decreasing the hydration of peptide backbone[4] based on an alanine rich peptide (AP) system. In our study, we come to the same conclusion as Asciutto et al. does by a hydration study and a PI energy calculation. The calculated Ψ-angle distribution of PLL in pure water does not agree with the CD experimental results. We calculated a large fraction of random coil structures while the experimental results showed the dominant structure for PLL in pure water is PPII structure.

The disagreement may be the result of using a set of force field parameters biasing α-helical structures. We will test this hypothesis by running a new set of metadynamics simulation using CHARMM36 force field parameters which corrects the biased potential parameters. In conclusion, our hypothesis that ClO_4^- changes the peptide backbone hydration and stabilizes α-helical structures is supported by the results from our enhanced MD simulations.

6.0 FUTURE WORK

The influence of different ions on PLL secondary structure can be examined by the free energy landscape of PLL, the Ψ-angle distribution of PLL, the hydration of the peptide backbone and the distribution of the ion-peptide interaction. In our work, we demonstrate that enhanced sampling MD method is useful in obtaining the conformational space of PLL in a short time. However, theΨ-angle distribution for PLL in pure water contradicted the CD experimental results. The discrepancy may be the result of using the CHARMM27 force field parameters in which the potential parameters for peptide backbone bias towards α-helical structures. To examine this hypothesis, get a better simulation result and extend the analysis method to other systems, we would like to focus the future work on the following two aspects:

1. Carry out the metadynamics simulation of PLL in salt solutions using a set of optimal parameters. For the force field parameters of atoms, we will use CHARMM36 force field which modified the CHARMM27 force field to correct the overweighting of α-helical structures[41] over PPII structures. Parameters of the history-dependent potential in metadynamics simulation will be determined in a similar fashion as in determining the parameters in the case of pentane simulation.
2. The effect of ions on hydrophobic segments in proteins is not easy to investigate by experiments because of the low solubility of hydrophobic peptides in solution. But the same problem is not encountered in computational simulations. In the future work, we would like to study the impact of biologically relevant ions on hydrophobic peptides as a model system to understand protein folding with hydrophobic segments. We would also like to extend the current salt systems to buffer solutions which contain ions that are

common in living creatures like phosphate.

BIBLIOGRAPHY

[1] Ma, L.; Hong, Z.; Sharma, B.; Asher, S. *J. Phys. Chem. B* **2012**, *116*, 1134–42.

[2] Ma, L.; Ahmed, Z.; Asher, S. *J. Phys. Chem. B* **2011**, *115*, 4251–8.

[3] Bello, J. *Biopolymers* **1992**, *32*, 185–188.

[4] Asciutto, E. K.; General, I. J.; Xiong, K.; Xiong, K.; Asher, S. a.; Madura, J. D. *Biophysical journal* **2010**, *98*, 186–96.

[5] Paterova, J.; Rembert, K. B.; Heyda, J.; Kurra, Y.; Okur, H. I.; Liu, W. R.; Hilty, C.; Cremer, P. S.; Jungwirth, P. *J. Phys. Chem. B* **2013**, *117*, 8150–8158.

[6] Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–52.

[7] Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput Mol Sci* **2011**, *1*, 826–843.

[8] Dellago, C.; Bolhuis, P. G. *Adv Polym Sci* **2009**, *221*, 167–233.

[9] Chipot, C. *WIREs Comput Mol Sci* **2014**, *4*, 71–89.

[10] Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714–21.

[11] Vorobyov, I.; Anisimov, V. M.; Greene, S.; Venable, R. M.; Moser, A.; Pastor, R. W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2007**, *3*, 1120–1133.

[12] Lee, H.; Venable, R. M.; Mackerell, A. D.; Pastor, R. W. *Biophys. J.* **2008**, *95*, 1590–9.

[13] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comp. Chem.* **2005**, *26*, 1781–802.

[14] Martin, J. M. L. *J. Phys. Chem. A* **2013**, *117*, 3118–32.

[15] Hofmeister, F. *Archiv fr experimentelle Pathologie und Pharmakologie* **1888**, *25*, 1–30.

[16] F. A. Long, W. F. M. *Chem. Rev.* **1952**, *51*, 119–169.

[17] Collins, K. D. *Biophysical chemistry* **2006**, *119*, 271–81.

[18] Heyda, J.; Lund, M.; Oncák, M.; Slavícek, P.; Jungwirth, P. *J. Phys. Chem. B* **2010**, *114*, 10843–52.

[19] Algaer, E. a.; van der Vegt, N. F. a. *J. Phys. Chem. B* **2011**, *115*, 1259–81.

[20] Lund, M.; Vacha, R.; Jungwirth, P. *Langmuir* **2008**, *24*, 3387–91.

[21] Rembert, K.; Paterova, J.; Heyda, J. *J. Am. Chem. Soc.* **2012**, *134*, 10039–10046.

[22] Ma, L.; Ahmed, Z.; Mikhonin, A. V.; Asher, S. A. *J. Phys. Chem. B* **2007**, 7675–7680.

[23] Henin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. *J. Chem. Theory Comput.* **2010**, *6*, 35–47.

[24] Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.

[25] Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.

[26] MacKerell, A.; Bashford, D. *J. Phys. Chem. B* **1998**, *5647*, 3586–3616.

[27] M. Baaden, F. Berny, C. Madic,; Wipff, G. *J. Phys. Chem. A* **2000**, 7659–7671.

[28] Joung, I. S.; Cheatham, T. E. *J. Phys. Chem. B* **2008**, *112*, 9020–41.

[29] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

[30] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

[31] Bruxelles, U. L. D. *J. Comp. Phys.* **1977**, *341*.

[32] Miyamoto, S.; Kollman, P. a. *J. Comp. Chem.* **1992**, *13*, 952–962.

[33] Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613.

[34] Brooks, B. R.; Iii, C. L. B.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M. *J. Comp. Chem.* **2009**,

[35] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187–217.

[36] MacKerel Jr., A.; Brooks III, C.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; v. R. Schleyer et al., P., Ed.; The Encyclopedia of Computational Chemistry; John Wiley & Sons: Chichester, 1998; Vol. 1; pp 271–277.

[37] Roe, D. R.; Cheatham, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

[38] R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2013.

[39] Tiffany, M. L.; Krimm, S. *Biopolymers* **1972**, *11*, 2309–16.

[40] Rucker, A.; Creamer, T. *Protein Sci.* **2002**, 980–985.

[41] Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; Mackerell, A. D. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.