2017

Investigating Secondary Structure Features of YAP1 Protein Fragments Using Molecular Dynamics (MD) and Steered Molecular Dynamics (SMD) Simulations

Ferdiemar Cardenas Guinto Jr.
University of the Pacific, f_guinto@u.pacific.edu

Follow this and additional works at: https://scholarlycommons.pacific.edu/uop_etds
Part of the Chemistry Commons, and the Pharmacy and Pharmaceutical Sciences Commons

Recommended Citation

This Thesis is brought to you for free and open access by the Graduate School at Scholarly Commons. It has been accepted for inclusion in University of the Pacific Theses and Dissertations by an authorized administrator of Scholarly Commons. For more information, please contact mgibney@pacific.edu.
INVESTIGATING SECONDARY STRUCTURE FEATURES OF YAP1 PROTEIN FRAGMENTS USING MOLECULAR DYNAMICS (MD) AND STEERED MOLECULAR DYNAMICS (SMD) SIMULATIONS

by

Ferdiemar C. Guinto Jr.

A Thesis Submitted to the
Office of Research and Graduate Studies
In Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE

Pharmaceutical and Chemical Sciences Program
Bioanalytical and Physical Chemistry

University of the Pacific
Stockton, California

2017
INVESTIGATING SECONDARY STRUCTURE FEATURES OF YAP1 PROTEIN FRAGMENTS USING MOLECULAR DYNAMICS (MD) AND STEERED MOLECULAR DYNAMICS (SMD) SIMULATIONS

by

Ferdiemar C. Guinto Jr.

APPROVED BY:

Dissertation Advisor: Charles Michael McCallum, Ph.D.

Committee Member: Jerry Tsai, Ph.D.

Committee Member: Doug Weiser, Ph.D.

Department Chair: Andreas Franz, Ph.D.

Interim Dean of Graduate Studies: James A. Uchizono, Pharm.D., Ph.D.
INVESTIGATING SECONDARY STRUCTURE FEATURES OF YAP1 PROTEIN FRAGMENTS USING MOLECULAR DYNAMICS (MD) AND STEERED MOLECULAR DYNAMICS (SMD) SIMULATIONS

Copyright 2017

by

Ferdiemar C. Guinto Jr.
DEDICATION

This thesis is dedicated to my friends and family, without whom, none of this would be possible. This has been a long journey from the start to my collegiate career to where I stand today, and I find it difficult to believe that I would be here without the support of everyone around me. I owe all of you a tremendous “thank you” and hope you all remain with me as I continue my journey and move forward. Special recognition goes to my Mother who has been extremely supportive throughout all of my endeavors. Everything I do, I do for all of you.
ACKNOWLEDGMENTS

I would like to acknowledge a handful of individuals, all of who contributed to my academic success in some way. Mr. Keith Fraga and Ms. Shavarni Patel; who entered the PCSP graduate program at Pacific the same time as me. Having two other students going through the motions of that difficult first semester together made life just that much more bearable. Dr. Hyun Joo and Mr. Wei Zheng; two incredible individuals and scientists, thank you for constantly being there to help in research and in life. I will always look back on the conversations we’ve had with admiration. Finally, but possibly most importantly, Dr. Silvio Rodriguez, Dr. Jerry Tsai, and Dr. C.M. McCallum; without these three individuals, I would literally not be where I am today. Thank you.
Investigating Secondary Structure Features of YAP1 Protein Fragments Using Molecular Dynamics (MD) and Steered Molecular Dynamics (SMD) Simulations (MD)

Abstract

by Ferdiemar C. Guinto Jr.

University of the Pacific
2017

Molecular dynamics (MD) is a powerful tool that can be applied to protein folding and protein structure. MD allows for the calculation of movement, and final position, of atoms in a biomolecule. These movements can be used to investigate the pathways that allow proteins to fold into energetically favorable structures. While MD is very useful, it still has its limitations. Most notable, computing power and time are of constant concern.

Protein structure is inherently important due to the direct link between the structure of a protein and its function. One of the four levels of protein structure, the secondary structure, is the first level to accommodate for the three-dimensional shape of a protein. The main driving force behind secondary structure is hydrogen bonding, which occurs between the carboxyl oxygen and the amine hydrogen of the backbone of a peptide. Determining a greater link between hydrogen bond patterns and types of secondary structure can provide more insight on how proteins fold.
Because molecular dynamics allows for an atomic level view of the dynamics behind protein folding/unfolding, it becomes very useful in observing the effects of particular hydrogen bond patterns on the folding pathway and final structure formed of a protein. Using molecular dynamic simulations, a series of experiments in an attempt to alter structure, hydrogen bonding, and folding patterns, can be performed. This information can be used to better understand the driving force of secondary structure, and use the knowledge gained to manipulate these simulations to force folding events, and with that, desired secondary structure features.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ 10

LIST OF FIGURES .................................................................................................... 11

CHAPTER

1. Introduction: Using Molecular Dynamics (MD) and Steered Molecular Dynamics (SMD) to Investigate Secondary Structure Properties of Yes-Associated Protein 1 (YAP1) Protein Fragments ................................................................................. 15

   1.1 Structure Determination: Experimental and Theoretical ................................ 20

   1.2 Computational Methods .................................................................................. 21

   1.3 MD Theory ........................................................................................................ 22

   1.4 Protein/Peptides of Interest .............................................................................. 22

   1.5 Experimental Methods and Materials ............................................................... 23

   1.6 Experimental Set Up ......................................................................................... 23

   1.7 Outline of Experimentation ............................................................................... 24

2. Molecular Dynamic Simulations on Peptide Fragments from the 1SSE Domain of the YAP1 Protein. Investigations on Secondary Structure Properties Through the Use of Solvent Effects, Dipole Manipulation, and Sequence Properties ........................................................................... 28

   2.1 Peptides of Interest and Preparation ............................................................... 33

   2.2 Preparing files .................................................................................................. 35

   2.3 Results and Discussion: .................................................................................... 38

   2.4 Peptide 301: Cysteine at the C-terminus ......................................................... 38

   2.5 Peptide 616: Cysteine in the Middle of the Peptide ......................................... 49

   2.6 Peptide 638: No Cysteine Present in the Peptide ............................................ 58
2.7 Peptide 598: Cysteine at the N-terminus of the Peptide ........................................ 67

2.8 Simulations of the Complete Domain ................................................................. 76

2.9 Conclusion/Explaining the Results ................................................................. 80

3. Using Molecular Dynamic Simulations on the 638 Peptide to Establish a Hierarchy of Secondary Structure Hydrogen Bond Patterns ........................................ 82

3.1 Abstract ............................................................................................................. 82

3.2 Introduction ....................................................................................................... 83

3.3 Methods and Materials ....................................................................................... 85

3.4 Experimental Setup .......................................................................................... 87

3.5 Results and Discussion ..................................................................................... 88

3.6 Hydrogen Bond Analysis .................................................................................. 98

3.7 Conclusion .......................................................................................................... 114

4. Steered Molecular Dynamic Simulations of the 638 Peptide Fragment to Illicit Unfolding Events in Order to Attempt to Better Understand the Folding Pathway .............................................................. 115

4.1 Abstract ............................................................................................................. 115

4.2 Introduction ....................................................................................................... 115

4.3 Methods and Materials ....................................................................................... 117

4.4 Experimental Setup .......................................................................................... 118

4.5 Results and Discussion ..................................................................................... 121

4.6 Folding Extended Peptides Using Standard MD .............................................. 124

4.7 Experimental Simulations: ............................................................................... 128

4.8 Determining Important Hydrogen Bonds ....................................................... 138

4.9 Conclusion .......................................................................................................... 146
5. An Attempt to Perform Peptide-Ion Interaction/SMD Simulations in the Aqueous Phase.......................................................................................................................... 147
   5.1 Introduction .................................................................................................................. 147
   5.2 Methods & Materials ................................................................................................... 147
   5.3 Results .......................................................................................................................... 148
   5.4 Interactions between +2 ions and peptide 638 in the gas phase ......................... 151
   5.5 SMD Simulations ......................................................................................................... 161
   5.6 Conclusions .................................................................................................................. 163
   5.7 Future Simulations ....................................................................................................... 163

6. REFERENCES ..................................................................................................................... 165

7. APPENDICES .................................................................................................................... 170
   7.1 Appendices: Chapter 1 ................................................................................................. 170
   7.2 Failed MD Simulations: ............................................................................................... 175
   7.3 Appendices: Chapter 3 ................................................................................................. 177
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Peptides of Interest from the YAP1 Protein</td>
<td>34</td>
</tr>
<tr>
<td>2. Active Hydrogen Bonds in the 638 Peptide MD Simulations</td>
<td>97</td>
</tr>
<tr>
<td>4. Quantifying Alpha-Helical Nature After the Removal of the Sodium Ion</td>
<td>107</td>
</tr>
<tr>
<td>5. Active Hydrogen Bonds in the 638 Peptide During SMD Simulations</td>
<td>135</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>3-D Representation of the 1SSE Domain</td>
<td>32</td>
</tr>
<tr>
<td>2.</td>
<td>Secondary Structure Timeline of the 301 Peptide (Gas Phase)</td>
<td>40</td>
</tr>
<tr>
<td>3.</td>
<td>Secondary Structure Timeline of the 301 Peptide (Sodium at CT)</td>
<td>42</td>
</tr>
<tr>
<td>4.</td>
<td>Secondary Structure Timeline of the 301 Peptide (Sodium at NT)</td>
<td>44</td>
</tr>
<tr>
<td>5.</td>
<td>Secondary Structure Timeline of the 301 Peptide (TIP3)</td>
<td>46</td>
</tr>
<tr>
<td>6.</td>
<td>RMSD for the 301 Peptide Simulations</td>
<td>48</td>
</tr>
<tr>
<td>7.</td>
<td>Secondary Structure Timeline of the 616 Peptide (Gas Phase)</td>
<td>50</td>
</tr>
<tr>
<td>8.</td>
<td>Secondary Structure Timeline of the 616 Peptide (Sodium at CT)</td>
<td>52</td>
</tr>
<tr>
<td>9.</td>
<td>Secondary Structure Timeline of the 616 Peptide (Sodium at NT)</td>
<td>54</td>
</tr>
<tr>
<td>10.</td>
<td>Secondary Structure Timeline of the 616 Peptide (TIP3)</td>
<td>56</td>
</tr>
<tr>
<td>11.</td>
<td>RMSD for the 616 Peptide Simulations</td>
<td>57</td>
</tr>
<tr>
<td>12.</td>
<td>Secondary Structure Timeline of the 638 Peptide (Gas Phase)</td>
<td>59</td>
</tr>
<tr>
<td>13.</td>
<td>Secondary Structure Timeline of the 638 Peptide (Sodium at CT)</td>
<td>61</td>
</tr>
<tr>
<td>14.</td>
<td>Secondary Structure Timeline of the 638 Peptide (Sodium at NT)</td>
<td>63</td>
</tr>
<tr>
<td>15.</td>
<td>Secondary Structure Timeline of the 638 Peptide (TIP3)</td>
<td>65</td>
</tr>
<tr>
<td>16.</td>
<td>RMSD for the 638 Peptide Simulations</td>
<td>66</td>
</tr>
<tr>
<td>17.</td>
<td>Secondary Structure Timeline of the 598 Peptide (Gas Phase)</td>
<td>68</td>
</tr>
</tbody>
</table>
18. Secondary Structure Timeline of the 598 Peptide (Sodium at CT) ..................... 70

19. Secondary Structure Timeline of the 598 Peptide (Sodium at NT) ................. 72

20. Secondary Structure Timeline of the 598 Peptide (TIP3) .......................... 74

21. RMSD for the 598 Peptide Simulations ..................................................... 75

22. Secondary Structure Timeline for Peptides of Interest (From the Domain) ...... 78

23. Secondary Structure Timeline for the 1SSE Domain .................................. 79

24. 3D Representation of the 638 Peptide Fragment ........................................ 86

25. Secondary Structure Timeline of 638 (Full Production MD) ....................... 89

26. Hydrogen Bond Timeline of 638 (Full Production MD) ............................. 90

27. Secondary Structure Timeline of 638 and SOD (Full Production MD) ......... 92

28. Hydrogen Bond Timeline of 638 and SOD (Full Production MD) ............... 93

29. 3-D Comparison of the Sodium’s Effect on Peptide 638 ............................ 95

30. Secondary Structure/Hydrogen Bond Timelines for Negative Control ......... 104

31. Timelines after the Removal of the Sodium Ion (Helix Stabilized) ............... 106

32. Quantifying Lifespans of Individual Hydrogen Bonds ............................... 109

33. Secondary Structure Timeline for SMD Simulation Without Equilibration .... 120

34. Secondary Structure Timeline for Peptide 638 Without Sodium Ion .......... 122

35. Secondary Structure Timeline for Peptide 638 With Sodium Ion ................. 123

36. Secondary Structure Timeline for Peptide 638 in an Attempted Folding ....... 125

37. Hydrogen Bond Lifespans During Attempted Folding Simulations ............... 127

38. Hydrogen Bond and Secondary Structure Timeline for Peptide 638 with Sodium SMD Simulations ................................................................. 130
39. Comparing the Hydrogen Bond Lifespans Between Successful and Failed Refolding simulations

40. Reverse Pulling Simulations to Compare Hydrogen Bond Breaking Patterns

41. Overlay of the Secondary Structure Timeline and Hydrogen Bond Timeline for the 15 Replicate SMD Simulations

42. Secondary Structure Timeline for Folding MD Simulations (Using Preserved Hydrogen Bonds)

43. Hydrogen Bond Lifespan for Hydrogen Bond Folding Simulations

44. Secondary Structure Timeline for Peptide 638 With and Without Sodium Ion

45. Secondary Structure Timeline for Peptide 638 With Potassium Ion

46. Secondary Structure Timeline for Peptide 638 With Magnesium Ion

47. 3-D Representation of Peptide 638 With Potassium Ion

48. Distance Between the C-terminus of the Peptide and the Sodium Ion

49. Secondary Structure Timeline of the 638 Peptide in the aqueous environments (with and without Mg ion)

50. Distance Between the C-terminus of the Peptide and the Mg Ion (AQ)

51. Secondary Structure Timeline for the 638 and Mg Ion SMD Simulations (AQ)
Chapter 1: Introduction: Using Molecular Dynamics (MD) and Steered Molecular Dynamics (SMD) to Investigate Secondary Structure Properties of Yes-Associated Protein 1 (YAP1) Protein Fragments

Abstract

Protein folding, structure, and function are major areas of interest of biophysical chemistry research. While the field is mature - with over 40 year of progress – protein folding is not fully understood.\textsuperscript{1,2} It has been very difficult to predict the corresponding secondary and tertiary structures from a protein’s primary amino acid sequence. Likewise, if a specific protein structure is desired, it has been very difficult to design a primary sequence, which will fold into the desired protein structure. The major barrier being that the complexity of protein folding cannot currently be well represented in a simple manner or visualized dynamically on the scale necessary to observe how and why the protein behaves the way it does during folding events. Computational chemistry allows for the evaluation and visualization of dynamical systems at up to an atomic level, lending itself to giving insight into the area of protein folding. In this thesis, secondary structure features of protein fragments from the Yes-Associated Protein 1 (YAP1) protein will be investigated and manipulated, in order to better understand the forces behind protein folding.
Introduction

Protein structure is vitally dependent in that the function of a protein is usually dependent on its structure. It has also been shown that many proteins favor specific structures based solely on the primary sequence. These native structures that proteins fold into are often reproducible. The main evidence for this has been attributed to Anfinsen’s experiment in which RNAse A was denatured and refolded to the native structure. Many attribute that the structures proteins fold into are derived from two main forces; hydrogen bonding for secondary structures and hydrophilic/hydrophobic interactions for tertiary structures. While hydrophilic/hydrophobic interactions are thought of as the main driving force for tertiary structures, hydrogen bonding is primarily responsible for all secondary structures. One way this manifests itself is that small peptides still have the capability to fold. As peptides contain less residues than proteins, hydrophobic packing becomes difficult, meaning that some other force must be responsible for protein folding besides hydrophobic packing.

Altering dipoles of peptides can have drastic affects on the stability of certain secondary structure features. By placing a cysteine at the C-terminus of a peptide, known as cysteine capping, has been shown both computationally and experimentally to stabilize the alpha helical nature of peptides. In addition, globular peptides have been influenced into alpha-helical structures by using ions in solution. Through the addition of charged species near globular proteins, it has been determined that the protein will take on a more helical structure, further displaying the affect of dipole manipulation on secondary structure.
Secondary structure is the first level of protein structure to account for the three-dimensional shapes of substructures of a protein. Molecular Dynamics (MD) is a very powerful tool in studying secondary structure. Its dynamic nature, as well as the ability to gather volumes of data, and the visualization of such small scales make it very probing in determining not only what is occurring during these protein folding events, but why certain features are more prominent than others. If hydrogen bonding is the main culprit in secondary structure folding events, MD should lend itself well to studying these causes and further explaining this driving force in protein folding.

Background

Protein structure/folding: importance and definition. Protein structure is important for a number of reasons; possibly most important, is the fact that the structure of a protein will dictate the function. It has been determined that two proteins that are ~80% identical in structure will display similar functionality.¹⁴

Proteins are relatively large polymeric biomolecules made up of building blocks known as amino acids, also referred to as residues. There are 20 naturally occurring amino acids in total, each of which have the same general backbone structure: a Nitrogen-Carbon-Carbon sequence makes up the backbone of every amino acid. The nitrogen in the amino acid sequence is the amine terminus (\(-\text{NH}_2\)) while the opposite end carbon is a carboxyl terminus (\(-\text{COOH}\)). The only differing quality amongst the separate amino acids is the side chain (R group) that is connected to the middle carbon.¹⁵ These amino acids are capable of forming peptide bonds, which result in the formation of peptides that can link into proteins.¹⁶ Peptide bonds occur when the N-
terminus nitrogen forms a bond with the C-terminus carbon of another amino acid. A condensation reaction takes place, resulting in the loss of water, and the formation of an amide. In order to distinguish the ends of a protein, the C-terminus and N-terminus nomenclature was developed. With a free carboxyl group existing at one end of the protein (the "C-terminus") and a free amino group at the opposite end (the "N-terminus"). The side-chains of these amino acids can be organized into separate groups; electrically charged side-chains (which can be positive or negatively charged), polar uncharged side-chains, non-polar (or hydrophobic) uncharged side-chains, and special cases. While most amino acids (with the exception of glycine) do have handedness, meaning they can form both L-stereoisomers ["left-handed" orientations] and D-stereoisomers ["right-handed" orientations]) in relation to the α-carbon, generally speaking we refer to the L-stereoisomers, as these are the most abundant isomers in nature.\textsuperscript{17,18}

There are four levels of protein structure: primary, secondary, tertiary, or quaternary. Primary structure is simply the sequence of the amino acids that make up the protein. The secondary structure is the first level of complexity to account for the three-dimensional structure of the amino acids, albeit at a local sub-structure level. In addition to the four separate levels of protein structure, there are also a handful of secondary structure types. This includes, but is not limited to, the alpha helix, the 3\textsubscript{10} helix, beta sheets, random coils, turns, etc., with each secondary structure being defined by the hydrogen bond patterns between the main-chain peptide groups.\textsuperscript{19} Tertiary structure refers to the three-dimensional structures formed between the packing of
separate secondary structures. This is the result of one continuous chain of amino acids forming different secondary structure features in separate regions of the sequence, and then folding together into shared spatial areas to form a stable structure. This level of folding is a result of hydrophobic interactions.\textsuperscript{3} Residues containing non-polar side chains (hydrophobic) will prefer to be away from any polar solvent, and as a result will be packed away into the structure. However, residues with polar side-chains will tolerate polar solvent, and thus will be exposed in the tertiary structure.\textsuperscript{20} Finally, quaternary structure is the result of multiple tertiary structures interacting together. This is the result of multiple amino acid chains folding into tertiary structures and fitting together.

Higher levels of protein structure are complicated, with multiple interactions needed to fold these multiple residue units into large three-dimensional structures. At the core, the structure of any protein originates from purely the primary sequence of the protein. There is information encoded into the sequence itself that determines how a protein will fold before any other interactions have a chance to take place. This idea is referred to as Anfinsen's Dogma\textsuperscript{6} and was a direct result of Anfinsen’s Experiment, which demonstrated the spontaneous folding nature of protein. Anfinsen used the protein RNase A as the protein of interest for his experiment. Through the use of β-mercaptoethanol and urea, he was able to denature the protein completely. The β-mercaptoethanol allowed the reduction of the disulfide bonds that stabilized the tertiary structure of the protein. The denatured protein was then completely dissolved in urea. Once fully dissolved, the urea was removed and the disulfide bonds re-oxidized,
resulting in renaturation of the protein. Among renaturation, the protein, the catalytic activity of RNase A was restored as well.6

**Structure Determination: Experimental and Theoretical**

There are many experimental techniques used to determine the structure and sequence of proteins. However, the most widely used method of protein structure determination is X-Ray crystallography. Crystallography allows for the determination of the three-dimensional protein structure at a specific stable state by shining x-ray beams onto a protein crystal structure. The atoms that make up the molecule will cause the x-ray beams to diffract and the diffraction of these beams can be observed and measured. Using the data gained from the diffraction a structure can be calculated and visualized.21,22 Another common technique used in the determination of protein structures is Nuclear Magnetic Resonance (NMR). NMR uses magnetic frequencies to hit a sample. The nuclei of the sample are affected and their spins are changed. Upon the spins returning back to their proper states, an electromagnetic frequency is re-emitted. This frequency can be recorded and used for determination of the position of specific atoms.21,23 Finally, Circular Dichroism (CD) is another technique in determining secondary structure attributes. CD makes use of polarized light. Due to the handedness of protein, specific states will interact only with specific yielding rotation of light. This information can be used to determine secondary structure features of a protein.24

While protein structure can be analyzed experimentally, each technique has it’s limitations when it comes to dynamics. Either the structures are static, with no indication of how the protein folded into structure (crystallography), an atomic level
representation is not achieved (CD), or size and time constraints become an issue (NMR). Computational chemistry lends a hand to this problem. With the limits of experimental techniques, computational chemistry can provide a dynamic approach to protein folding, demonstrating a "how" or "why" proteins fold the way they do, at an atomic scale, filling this void.

**Computational Methods**

Molecular Dynamics (MD) is a form of theoretical computational chemistry in which the movement and position of atoms in a molecule at a molecular level are predicted and simulated. One of the biggest advantages to MD is the fact that the results provide a dynamical approach, and representation, to macromolecular systems. MD simulation programs utilize force fields created through experimental and quantum mechanical approaches in order to describe how different atoms should interact alone, and with one another.\(^{25}\) The main goal of an MD simulation is to provide structures, which are energetically favorable and stable. Using Laws of Thermodynamics and Newton’s Equations of Motions, the movement of atoms can be predicted to create a dynamic representation of protein folding.

Steered Molecular Dynamics (SMD) is a subset of MD. While MD allows the atoms of the system to move freely, SMD gives the user the option of applying forces to specific atoms within the system.\(^{26,27}\) These forces can be used to either hold an atom in place, or cause an atom to move, which is useful in initiating or controlling motion, such as directed folding simulations. Because SMD allows for a detailed look at
unfolding events, it is a very useful tool for investigations regarding protein folding/unfolding.

**MD Theory**

MD programs calculate the positions of all atoms - for coarse grain models - at every time step of the simulation using a hard sphere approximation for the atoms.\(^{28}\) It is common for individual time steps to last 2 femtosecond. This is not an arbitrary value as the time it takes for a bond between to atoms to stretch is approximately 1 fs. Meaning that a 2 fs time step can account for a bond stretching and returning. While it is possible to calculate position and energy at every time step, this can become very costly computationally. In most cases, molecular dynamics simulations in the nanosecond scale range are wanted. This contributes enough time for the molecule to fold into alpha-helical secondary structures. This results in millions of steps and can take time and effort to calculate as well as analyze. This is one of the biggest drawbacks of computational chemistry, and specifically molecular dynamics.

**Protein/Peptides of Interest**

The YAP1 is an important protein thought to be an oncogene.\(^{29-31}\) Because of its biological importance, much experimental research has been performed on it, and structural data has already been determined.\(^{32,33}\)

In addition to its biological importance and known structure, the components of the protein also make it a protein of interest for molecular dynamics simulations. The YAP1 protein contains a domain (1SSE), which contains four separate helices, each with different placement of a cysteine residue; one with a cysteine at the C-terminus, one
with the cysteine at the N-Terminus, one with a cysteine in the middle of the peptide, and one with no cysteine present. Thus, this one domain of the YAP1 Protein can provide a somewhat complete range of cysteine positions for study.  

**Experimental Methods and Materials**

A series of computational programs are used in the process of setting up simulations, performing simulations, analyzing the dynamics, and analysis of quantifiable data. Nanoscale Molecular Dynamics Program (NAMD)\textsuperscript{35} is the program used to perform all molecular dynamic simulations. Visual Molecular Dynamics (VMD)\textsuperscript{36} is used to analyze the binary trajectory files (DCD files) as well as prepare systems (solvation, ionization, and orientation) through the use of CHARMM/MMTSB\textsuperscript{37,38} commands. Finally, the use of coding/graphing programs is used to display quantitative data in an appropriate manner.

**Experimental Set Up**

A variety of setups are utilized to investigate many different properties, interactions, structures, and forces. Different environments can be utilized, such as vacuum or solvated environments (which is in most cases a water solvent). Ions can be placed into the system and at varying locations. Different types of ions, or both positive and negative charge, can also be used. The results of these varying setups can be compared to gain further insight as to why and how these interactions are taking place. For example, simulations of the peptide of interest only and simulations of the peptide of interest plus the addition of a sodium ion can be compared to determine the effects of the ion. Sodium ion has been shown to stabilize helix experimentally.\textsuperscript{39}
Due to the natural limitations in computational power and time constraints, preliminary simulations were first performed on short time scales. These short times ranged from 200 picoseconds-1000 picoseconds (0.2 ns to 1ns, respectively), which allowed the investigation of results without the risk of too much time investment. If the results were determined to be useful, then much longer timescales could be utilized for further investigation.

Outline of Experimentation

Molecular dynamic simulations on peptide fragments from the 1SSE Domain of the YAP1 Protein. Investigations on secondary structure properties through the use of solvent effects, dipole manipulation, and sequence properties. Initially investigations focused on the secondary structure of peptide fragments in different environments. The secondary structures of the fragments in the gas phase, as well as the aqueous phase, were compared. After this, the effect of ion placement was investigated.

Each peptide of interest underwent four main types of simulations directed toward the area of study. The peptide by itself was simulated in a gas phase environment, the peptide by itself was simulated in a aqueous phase environment, the peptide and a sodium ion at the C-terminus are simulated in a gas phase environment, and the peptide and a sodium ion at the N-terminus are simulated in a gas phase environment. These four types of simulations provide coverage for secondary structure analysis to determine the effect of a sodium ion, as well as the effect of a solvent, will have on the corresponding secondary structure of the peptide. The peptide alone in a gas phase environment provides a control for the three other simulation types.
The four peptides were taken and simulations under a variety of simulations were performed in order to compare not only between individual peptides, but across the different peptides as well. Peptide 301 (which had the cysteine at the C-terminus), Peptide 616 (which had the cysteine in the middle of the peptide), Peptide 638 (which had no cysteine), and Peptide 598 (which had a cysteine at the N-terminus, as well as one cysteine near the C-terminus) were all used for experimental simulations. Initially, the first peptide chosen for experimentation was peptide 301. This is the alpha helix from the 1SSE domain that had a cysteine residue located at the C-terminus of the peptide. As with all experiments, the native structure must be determined through MD simulations. The peptide was placed in the vacuum environment and MD simulations were performed. Next, the structure of the peptide in an aqueous environment was determined through the addition of a TIP3 water box. Finally, the effect of a charged species on the structure of the peptide was investigated. For this, a sodium ion was placed at both the C-terminus and N-terminus to determine any structural changes. The other 3 peptides (616, 638, and 598) all underwent this same process.

**Conclusions.** Multiple conclusions can be drawn from the simulations comparing the structure of the peptide with different cysteine placements in different environments. It can be established that the presence of a positively charged ion at the C-terminus of a peptide can result in the stabilization of alpha helical nature if the peptide is absent of a cysteine residue or if the cysteine residue is located at the C-terminus of the peptide. If the cysteine is located at the N-terminus or in the direct middle of the peptide, the results were inconclusive. Using gas phase and aqueous
phase environments allows the determination of a peptide is hydrophobic or hydrophilic overall in nature. This is accomplished by using alpha helicity in differing environments as a gauge for stability.

**Long molecular dynamic simulations of Peptide 638 (no cysteine) to determine detailed secondary structure investigations.** Due to the fact that the peptide fragment 638 contains no cysteine, this makes it a very useful fragment for control simulations. This allowed experimentation of a much longer timescale than previously performed. Simulations with these fragments were increased to 10,000ps-100,000ps (10ns-100ns) simulations. Because 638 lacked cysteine residues in the sequence, it made it the perfect candidate for more in depth investigations. The lack of cysteine resulted in a lack of extraneous interactions between the cysteine residues and the ions, resulting in more expected secondary structure results. Using the previous simulations shows what is happening structurally, longer simulations can show why these structural features form, leading to how hydrogen bonds are key component to secondary structure.

**Peptide 638 SMD simulations: investigating folding events.** After using MD simulations to determine structural changes as well as the resulting changes in hydrogen bond patterns, this information was used in an attempt to stimulate protein folding. After SMD simulations, coupled with the information gained about the “importance of specific hydrogen bonds” using the MD data, a hierarchy of hydrogen bonds was determined. This hierarchy was used to fold peptides to more native structures than standard MD simulations typically allows.
**Aqueous simulations.** While it is very important to understand the intrinsic properties of these peptides, their structural characteristics, and their folding behavior, it is also important to study them in more natural environments. For this, a transition to aqueous phase environments is important. Most experimental parameters from the gas phase simulations were retained, (i.e., length of simulations, number of steps/output, temperature, volume, etc.) with the only difference being added solvent. Multiple factors on structure were once again compared. The structure of the peptide by itself in the gas phase versus the peptide by itself in the aqueous phase was compared once again, but this time at longer time scales. This allowed the visualization of solvent effects on the peptide in more detail. A sodium ion was added near the C-terminus of the peptide in both environments. What was observed was the difference in effect of a +1 charged ion on the peptide in the gas phase and the aqueous phase. As opposed to the interaction that was observed between the peptide and the sodium ion in the gas phase, the peptide and sodium ion do not seem to interact in aqueous phase. Instead, the ion floats away from the peptide, indicating that the solvent effect is too strong for neutral peptides. The interaction of the 638 peptide and a +2 ion (Magnesium) was then investigated in the gas phase. However, the ion in this case was too charged, forcing the entire peptide to interact with the ion, resulting in a “solvating” effect of the peptide around the ion. With these two examples in mind - the +1 ion is too weak in the aqueous phase and the +2 ion is too strong in the gas phase – an attempt to use the +2 ion was used in the aqueous phase. This provides observable results.
Chapter 2: Molecular Dynamic Simulations on Peptide Fragments from the 1SSE Domain of the YAP1 Protein. Investigations on Secondary Structure Properties Through the Use of Solvent Effects, Dipole Manipulation, and Sequence Properties

Introduction

Cysteine is a unique amino acid. As one of only two of the 20 total naturally occurring amino acids that contain sulfur (the other being methionine), its uniqueness is readily apparent in its chemical composition, but is further seen in its properties and activity. The reactivity of cysteine differs from methionine due to the fact that cysteine contains a sulfhydryl group - a single hydrogen is bonded directly to sulfur - whereas methionine contains sulfur bonded to a methane group. This difference in connective groups results in vast differences in the acidity, which in turn, directly influences the difference in reactivity between the two amino acids. The readiness of the cysteine residue to lose the hydrogen of the sulfhydryl group allows for two cysteine residues to easily form disulfide bridges, which have drastic effects on the secondary, tertiary, and quaternary structures of protein and their stabilities.

Cysteine placement has been shown to have an effect on the secondary structure of small peptides.\textsuperscript{12,34,40} Depending on where a cysteine is present in the sequence of a peptide can have an effect on the three-dimensional shape the peptide forms. The YAP1 protein is a naturally occurring, biologically active protein, which
contains a specific domain, the 1SSE domain. This domain contains four alpha helices, each with a differing placement of a cysteine residue in the sequence of the helix. The fact that the 1SSE domain contains four naturally occurring helices with cysteine in different locations makes the 1SSE domain a prime example to investigate the effect of cysteine placement on the secondary structure of alpha helical peptide fragments.

In addition to cysteine placement, it is also known that the environment that a protein is placed in will also have an effect on structure. Proteins are prone to solvent effects due to hydrophobic packing. Due to the fact that "like dissolves like", chemicals and compounds would rather be contained in environments that have similar properties. Polar groups would prefer to be in areas of other polar groups and non-polar groups like to be around areas of other non-polar groups. This results in a protein packing in such a manner that will optimize this relationship.

While initial investigations focused on the effect of cysteine placement on secondary structure features, additional investigations focused on the secondary structure features of the peptide fragments in varying environments. The secondary structures of the fragments in the gas phase, as well as the aqueous phase, were compared. Finally, the different effects of the placement of a charged species near these peptide fragments were investigated.
Methods/Materials

**Background.** To begin with, it is important to determine if the environment is properly set up, the simulations are equilibrated, so that the results are as understandable and can be interpreted. Because computational power and cost (in the form of time) are limited, "shorter scale" preliminary simulations must first be performed to determine if any necessary alterations will need to be made to the simulations before "experimental (longer timescale)" simulations are to be performed.

**Protein of interest.** The 1SSE domain of the YAP1 protein provides a useful subject for the simulation experiments. A multitude of reasons arise for using the YAP1 protein, and specifically the 1SSE domain, as the basis for this study. As mentioned, cysteine placement can be an important factor among peptide/protein secondary structure studies. The fact that the 1SSE domain contains multiple helices, all of which have a cysteine at differing positions, make this an attractive candidate for these computational experiments. In addition to this, the YAP1 protein is a biologically active molecule and known oncogene, as it has been seen to be overexpressed in cells that develop cancerous behavior, thus a better understanding of YAP1 may have an as of yet unrealized benefits resulting from the computational studies. Finally, because it is biologically important, there are many background experimental results that can be retrieved. In addition to the function, the structure of the protein has also already been solved. The amount of information already known about the protein, the different experimental results, combined with the still unknown features prove this protein to be a very useful subject of investigation with many different areas to delve into.
The most important factor in being able to use the YAP1 protein for theoretically investigation is that the whole of the YAP1 protein structure has been determined experimentally - however in parts - as seen by the structures of the separate domains. The use of Nuclear Magnetic Resonance (NMR) and protein crystallography studies provide two experimentally techniques to provide the structural data for this protein. This experimental data can be used to create structure files for the YAP1 protein than can be retrieved from the RCSB Protein Data Bank. Thus, starting structures of the YAP1 Protein fragments to be used during MD simulations are readily produced.
Figure 1: 3-D Representation of the 1SSE Domain
A three-dimensional representation of the 1SSE Domain from the YAP1 Protein. The entire protein is color-coded based on secondary structure features. The four peptides of interest are also color coded based on secondary structure and the alpha helical nature is visible. The α-carbons of the N-terminus and C-terminus for each peptide of interest are labeled with a VDW representation for the α-carbon itself, as well as a label. All cysteine residues are labeled by color-coding residue type (cysteine is yellow) as well as the atoms of the cysteine residues are placed in CPK representations.
Peptides of Interest and Preparation

**Obtaining peptides of interest.** Because the investigation is to be performed at the secondary level of protein structure, it is first necessary to isolate the peptides of interest from the larger protein domain. The protein domain PDB structure file - obtained from the RCSB Protein DataBank - is used to generate all subsequent peptide coordinate/structure files to be used for simulations.

While the structure file for the 1SSE Domain has been solved experimentally, there are no experimental results for the peptides of interest themselves. Because of this, computation simulations must be utilized in order to determine "the native structures" - which will be considered to be reference structures - of just the peptides of interest.
<table>
<thead>
<tr>
<th>Peptide Reference Code</th>
<th>Peptide Number Sequence</th>
<th>Peptide Amino Acid Sequence (NT to CT)</th>
<th>Number of Residues</th>
<th>Location of Cysteine in Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>301</td>
<td>301-310</td>
<td>EFCSKMNQVC</td>
<td>10 Residues</td>
<td>C-terminus</td>
</tr>
<tr>
<td>598</td>
<td>598-607</td>
<td>CSEIWRDRTT</td>
<td>10 Residues</td>
<td>N-Terminus</td>
</tr>
<tr>
<td>616</td>
<td>616-624</td>
<td>VDGLCSELM</td>
<td>9 Residues</td>
<td>Cysteine In the middle of the peptide</td>
</tr>
<tr>
<td>638</td>
<td>637-648</td>
<td>NAEDVQLALNKH</td>
<td>12 Residues</td>
<td>No Cysteine</td>
</tr>
</tbody>
</table>
Preparing files

Preparing specific peptides and desired environments. Each peptide undergoes a series of preparation steps in order to create the proper files necessary in order to perform molecular dynamic simulations. First, the entire protein domain is loaded into the Visualizing Molecular Dynamics (VMD) program. The sequence of amino acids corresponding to each peptide of interest is used to create clean PSF/PDB file combinations of each peptide of interest. The residue chain of the desired peptide is set and the Peptide files are generated using PSFGEN commands - which utilizes writepsf and writepdb commands. This results in final, and clean, coordinate and structure file pairs of only the peptide of interest.

The terminus of the peptide is capped by applying patches to each terminus. This provides a charge less group to avoid. By using the first ACE (acetylated NT), last CT3 (CT, N-methylamidated) commands, the desired caps are added to the peptide.

This new PDB/PSF file combination is loaded into VMD once more in order to create the separate environments that the different types of simulation will be performed in. The peptide is oriented around the Z-axis to begin with. The resulting aligned structures are used further to create systems of varying environments of interest.

Steps for creating specific environments: solvating/ionizing systems. The setups for the experimental simulations all undergo the same steps of preparation. First, the addition of the sodium ion must be accomplished. This is the result of a number of steps before leading to the final PSF/PDB combination. First, the clean, aligned, peptide
structure/coordinate files are solvated using VMD’s Solvation Modeling Extension plug-in. Once the solvated system is created, the system is ionized in a very similar manner to the solvation. The VMD Ionization Modeling Extension plug-in is utilized for this step. Once the solvated and ionized system is created, then the water must be removed. With the removal of the water, only the peptide and the sodium ion remain, and a clean PSF/PDB combination are created of the peptide + ion system. The ion is then moved around to the desired locations. First, the ion is moved to the C-terminus of the peptide. Once at the exact desired location, the PSF/PDB is created. The ion is then again moved, but this time, near the N-terminus of the peptide. With the creation of the PSF/PDB files of the peptide and the ion - at both desired positions - is created, then the PSF/PDB of the peptide without ion is saved. All necessary PDB/PSF combination files for the gas phase simulations are created at this point and ready for simulation.

For the aqueous phase simulations, the solvation package must be utilized. Solvation: aligned psf/pdb pairs are solvated directly. New psf/pdb pair of solvated system (tip3) saved for aqueous simulations.

**NAMD configuration files/experimental parameters.** While the total length of the simulations is conserved through the equilibration and MD simulations, other time factors are also held constant. Each timestep lasts 2 femtoseconds with a frame being produced every 500 steps. While it may be useful to produce a frame of data for each timestep, it is not necessary due to the time needed for atom/bond motions to occur. Limiting the amount of frames produced limits the file size of the output file, saving time, space, and allowing for easier analysis. The temperature of the system is set to
310 Kelvin, which remains constant, and the pressure set to 1 atmosphere. Once the configuration files are prepared, they can be used for future simulations with only minor alterations.

**Experimental set-up.** Due to the limitations of computational power it is first reasonable to perform all simulations on "shorter timescales" than experimental simulations. Because simulations can be very time expensive, these shorter simulations can be performed to provide a baseline of occurrences during the simulation of these peptides. For these reasons, the preliminary simulations are performed on the 200-500 ps timescales.

**Equilibration parameters.** Equilibration simulations for the peptides of interest are performed over 600 ps. This includes a 100 ps range of minimization and an additional 500 ps range of equilibration.

**MD simulation parameters.** Experimental MD simulation parameters are similar to equilibration parameters. The length of the simulation is lowered to 500 ps, as the 100 ps minimization steps are removed. The equilibrated structures from the equilibration simulations are used as the starting coordinates for all further experimental MD simulations.

**Computational programs.** All simulations are performed in NAMD (NAnoscale Molecular Dynamics Program) using CHARMM (CHemistry at HARvard Molecular Modeling) force fields. CHARMM22 is used for proteins and CHARMM27 for nucleic acids. Visualization of the trajectory files utilized VMD (Visualizing Molecular Dynamics Program) with additional analysis utilizing plug-ins of the VMD Program.
Results and Discussion

Cross-analysis experiments were performed for all four peptides of interest with different factors and environments being applied to each peptide. Both the different characteristics of the peptides and the effect of the environment on the individual peptides were investigated. Finally, the effect of the environments on each of the peptides could be compared to each other, once all other "baseline" simulations are finalized, providing substantial data.

The four peptides of interest are used for the following investigations:

1) Secondary structure attributes in a gas phase environment
2) Secondary structure attributes in an aqueous phase environment
3) Secondary structure attributes with charge species placed nearby (at both the C-terminus and N-terminus)

Peptide 301: Cysteine at the C-terminus

The first peptide used is referred to as peptide 301. This peptide contains a cysteine residue directly at the C-terminus. The peptide is taken and placed in a gas phase environment with no other compounds. The native secondary structure is the first desired piece of data. The secondary structure timeline displays a peptide without a very strong alpha helical nature throughout the peptide during the entire length of the simulation (equilibration and MD). Using VMD’s secondary structure timeline plug-in, the changes of the secondary structure throughout the simulation can be displayed in a two-dimensional graph. VMD will attribute a secondary structure type to each residue of the peptide for every step of the simulation. The secondary structure timeline will
then color code this secondary structure type and graphically show these changes.
Areas that are pink display the alpha helical nature that is being investigated. Areas of blue are representative of 310 helix, while red is indicative of $\pi$-helix. Finally, areas that are green and areas that are white display turns/coil and no secondary structure features, respectively. VMD uses phi and psi angles to make these determinations.
Figure 2: Secondary Structure Timeline of the 301 Peptide (Gas Phase)
The Secondary Structure Timeline for the 301 Peptide Fragment only in a gas phase environment (equilibration (top) and MD simulation (bottom))
Next, the effect of an associated ion was determined. As previously mentioned, the orientation of the peptide is maintained, and the gas phase environment is unchanged, however a sodium ion is placed near the C-terminus of the peptide. Because the peptide already displayed alpha helical nature without the presence of an ion, not much change in structure was expected; the resulting secondary structure timeline confirmed this idea.
Figure 3: Secondary Structure Timeline of the 301 Peptide (Sodium at CT)

The Secondary Structure Timeline for the 301 Peptide Fragment with a sodium ion at the C-terminus of the peptide. In a gas phase environment (equilibration (top) and MD simulation (bottom)).
While it was expected that the sodium ion would stabilize the alpha helical nature, that effect cannot be confirmed at this point because of the lack of change in the secondary structure features between the two simulations. There seems to be some stabilization at the N-terminus of the peptide.

In order to investigate any possible destructive ion-dipole behavior, the sodium ion is moved from the C-terminus of the peptide to the N-terminus. This resulted in very unexpected behavior from what was initially expected to occur. Analysis of the secondary structure timeline reveals the retention of a relatively alpha helical structure. This was unexpected, as the placement of the sodium ion at the N-terminus was predicted to have a destabilizing effect on the helix.
Figure 4: Secondary Structure Timeline of the 301 Peptide (Sodium at NT)
The Secondary Structure Timeline for the 301 Peptide Fragment with a sodium ion at the N-terminus of the peptide. In a gas phase environment (equilibration (top) and MD simulation (bottom))
Finally, the peptide was simulated in the aqueous phase for comparison to the gas phase simulations. While the peptide displayed strong alpha helical nature in all experimental simulations, the aqueous phase secondary structure results lined up best with the results obtained from placing the sodium ion at the C-terminus of the peptide.
Figure 5: Secondary Structure Timeline of the 301 Peptide (TIP3)
The Secondary Structure Timeline for the 301 Peptide Fragment only in an aqueous phase environment (equilibration (top) and MD simulation (bottom))
Due to the lack of visual changes in the secondary structure timelines, RMSD analysis was used to determine how much, if any, the effect of the ion and solvent have on the secondary structure of the peptide. A pure alpha helix poly alanine peptide is used as a starting structure to compare the RMSD time series of each peptide. Here, some discrimination can be made between the effects of the associated ion on the secondary structure of the peptide. While the peptide with the sodium ion at the C-terminus has a larger average RMSD, the standard deviation is much smaller, showing the stabilizing effect of the ion.
Figure 6: RMSD for the 301 Peptide Simulations
RMSD Time series for 301 Peptide in different environments
**Peptide 616: Cysteine in the Middle of the Peptide**

As before, the 616 peptide undergoes a series of experimental MD simulations to better understand the secondary structure features under different conditions.

Again, a standard place to begin is with the native structure of the peptide alone in a gas phase environment. The resulting secondary structure timeline displays a very disordered peptide. The only secondary structure features are turns. It can be determined that this helix requires an outside influence (the tertiary structure of the rest of the protein, water/solvent effects, etc.) in order to retain its helical nature that occurs in the native protein.
Figure 7: Secondary Structure Timeline of the 616 Peptide (Gas Phase)
The Secondary Structure Timeline for the 616 Peptide Fragment only in a gas phase environment (equilibration (top) and MD simulation (bottom))
Next, the effect of the placement of a sodium ion at the C-terminus is tested. The results from this type of simulation were once again unexpected prior to the final investigation. It was expected that the addition of a positively charged species at the C-terminus of a peptide would result in a formation/stabilization of alpha helical nature in a peptide. But once again, the peptide displayed very disordered properties. While there is an increase in "secondary structure" (turn) and a sparse amount of 310 helix, it is still very disordered throughout.
Figure 8: Secondary Structure Timeline of the 616 Peptide (Sodium at CT)
The Secondary Structure Timeline for the 616 Peptide Fragment with a sodium ion at the C-terminus in a gas phase environment (equilibration (top) and MD simulation (bottom))
While it seems that the sodium ion will have a destructive interaction with the 616 peptide fragment no matter where it is placed, the sodium ion was also placed near the N-terminus for due diligence. As expected, the ion and the cysteine in the middle of the peptide seem to interact, and the secondary structure of the peptide is destroyed.
Figure 9: Secondary Structure Timeline of the 616 Peptide (Sodium at NT)
The Secondary Structure Timeline for the 616 Peptide Fragment with a sodium ion at the N-terminus in a gas phase environment (equilibration (top) and MD simulation (bottom))
Finally, the 616 peptide is simulated in the aqueous phase. While the peptide struggles to form any sort of alpha helical structure during all simulations of the gas phase, the alpha helical features formed much more easily during the aqueous phase simulation. This would indicate that the peptide is completely hydrophilic in nature overall.
Figure 10: Secondary Structure Timeline of the 616 Peptide (TIP3)
The Secondary Structure Timeline for the 616 Peptide Fragment only in a aqueous phase environment (equilibration (top) and MD simulation (bottom))
Figure 11: RMSD for the 616 Peptide Simulations
RMSD Timeseries for the 616 Peptide Fragments in Various Environments
The RMSD timeseries displays some features of the stability of the alpha helical nature. The vacuum simulation of the peptide only has the highest RMSD with a large amount of fluctuations, showing very little stability. The simulation in explicit water solvent has a very low average RMSD, but also has a seemingly high standard deviation. Showing that while the structure is mostly stable overall, there seems to be points in the simulation that are unstable, which may be attributed to folding events. Finally, both simulations with the sodium ion in the gas phase have very small standard deviations, and relatively good average RMSDs. The most surprising piece of data however, is the small standard deviation of the ion placed near the N-terminus of the peptide, as this was expected to produce a destabilized structure.

**Peptide 638: No Cysteine Present in the Peptide**

Due to the unexpected results of the simulations utilizing peptide 301 and peptide 616, a transition to a "control" peptide was expected to clarify what was obtained. As peptide 638 contains no cysteine residues present in the sequence, it was determined simulations with this peptide would provide simplified results. Peptide 638 was placed in a gas phase environment with no other species in order to obtain a “native” structure. What resulted was a peptide with a reasonable portion of alpha helical nature, but enough variability to determine that the alpha helical structure was not "fully stabilized" – e.g., a fair amount of statistical noise is present in the secondary structure timeline for peptide 638.
Figure 12: Secondary Structure Timeline of the 638 Peptide (Gas Phase)
The Secondary Structure Timeline for the 638 Peptide Fragment only in a gas phase environment (equilibration (top) and MD simulation (bottom))
Next, the positively charged sodium ion was placed near the C-terminus of the peptide. The resulting secondary structure changes were along the lines of expected results. The alpha helical nature is stabilized and undergoes a shift from the N-terminus of the peptide to the C-terminus of the peptide; indicating a direct influence of the sodium ion on the peptide. Thus, the structure is not only stabilized, but part of the peptide is completely altered as well.
Figure 13: Secondary Structure Timeline of the 638 Peptide (Sodium at CT)
The Secondary Structure Timeline for the 638 Peptide Fragment with a sodium ion at the C-terminus in a gas phase environment (equilibration (top) and MD simulation (bottom))
Once again, in order to investigate possible destructive ion-dipole effects, a sodium ion was moved from the C-terminus to the N-terminus. The results corresponded well with what was expected prior to the simulation. The sodium ion no longer has a constructive, stabilizing effect, but rather a destructive effect. The alpha helical nature that was seen in simulations with the sodium ion at the C-terminus, as well as the simulations of just peptide 638, is completely removed.
Figure 14: Secondary Structure Timeline of the 638 Peptide (Sodium at NT)
The Secondary Structure Timeline for the 638 Peptide Fragment with a sodium ion at the N-terminus in a gas phase environment (equilibration (top) and MD simulation (bottom))
Finally, aqueous simulations of the 638 peptide were performed in order to display the solvent effects on the structure of the peptide. Although the peptide alone in the gas phase simulations proved to be fairly alpha helical, the aqueous simulations displayed much more alpha helicity.
Figure 15: Secondary Structure Timeline of the 638 Peptide (TIP3)
The Secondary Structure Timeline for the 638 Peptide Fragment only in an aqueous phase environment (equilibration (top) and MD simulation (bottom))
Figure 16: RMSD for the 638 Peptide Simulations
RMS Timeseries for the 638 Peptide Fragment in Various Environments
Peptide 598: Cysteine at the N-terminus of the Peptide

The final peptide to be used for simulations is peptide 598. This peptide contains a cysteine exactly at the N-terminus. To determine a native structure, peptide 598 alone was simulated in a gas phase environment. The resulting structure is very disordered throughout the majority of the simulation. However, some alpha helical nature does seem to form toward the tail end of the simulation.
Figure 17: Secondary Structure Timeline of the 598 Peptide (Gas Phase)
The Secondary Structure Timeline for the 598 Peptide Fragment only in a gas phase environment (equilibration (top) and MD simulation (bottom))
Next, the 598 peptide is simulated with a sodium ion placed near the C-terminus. A very unexpected timeline is produced. The majority of the peptide is structured with pi-helix attributes.
Figure 18: Secondary Structure Timeline of the 598 Peptide (Sodium at CT)
The Secondary Structure Timeline for the 598 Peptide Fragment with a sodium ion at the C-terminus in a gas phase environment (equilibration (top) and MD simulation (bottom))
The 598 peptide is simulated with a sodium ion placed near the N-terminus. In this case, the structure of the peptide once again seems to be disrupted as the resulting secondary structure timeline results in a mostly coil/turn representation.
Figure 19: Secondary Structure Timeline of the 598 Peptide (Sodium at NT)
The Secondary Structure Timeline for the 598 Peptide Fragment with a sodium ion at the N-terminus in a gas phase environment (equilibration (top) and MD simulation (bottom))
Finally, the 598 peptide alone is simulated in an aqueous environment. Like previous peptides that seemed hydrophilic, the gas phase simulations produced disordered secondary structures, however, once added to the aqueous phase environments, the alpha helical nature becomes very prominent.
Figure 20: Secondary Structure Timeline of the 598 Peptide (TIP3)
The Secondary Structure Timeline for the 598 Peptide Fragment only in an aqueous phase environment (equilibration (top) and MD simulation (bottom))
Figure 21: RMSD for the 598 Peptide Simulations
RMSD Timeseries for the 598 Peptide in Various Environments
Simulations of the Complete Domain

In order to and assess how the 1SSE reaction center domain interacts structurally, the whole of the domain is used for simulations. With the “native” structures of all four peptides from the 1SSE domain determined using MD simulations, a comparison with the naturally structures is sought. Due to the large size of the domain, the simulation parameters were slightly altered. The full length of the MD simulations was scaled back from 500 ps to 200 ps.
Figure 22: Secondary Structure Timeline for Peptides of Interest (From the Domain)  
The Secondary Structure Timeline for the four Peptide Fragments isolated from the simulations of the entire domain (301 (1st), 598 (2nd), 616 (3rd), 638 (4th)).
Figure 23: Secondary Structure Timeline for the 1SSE Domain
The Secondary Structure Timeline for the Whole 1SSE Domain
As expected, the four peptides of interest present very strong alpha helical nature when part of the entire domain. This would suggest that while tertiary folding has an effect on the secondary structure - the peptides form helices as a whole protein - there are also forces that are applicable to secondary structure formation - different structures occur in the peptides when removed from the entire domain.

**Conclusion/Explaining the Results**

With all four peptides simulated in both gas phase and aqueous environments, as well as with and without the presence of ions at different locations, the effect of these different factors on the secondary structure of the peptides can be further investigated. Initially, it was expected that the positive charge of the sodium ion placed at the C-terminus would stabilize any helical structure and the positive charge of the sodium ion placed at the N-terminus of the peptide would destabilize any alpha helical structure of that peptide, but what was observed seemed far more interesting. While this idea held true for the peptide containing no cysteine residues, all other peptide (which contained cysteine residues in the sequence) reacted very differently than expected. It seems that the cysteine residue is causing different interactions with ions. The cysteine at the C-terminus preserved the ability of the ion to stabilize an ordered helix, it also allowed the ion to preserve the helical nature if placed at the N-terminus as well. However, if there is a cysteine residue located at only the N-terminus, but not the C-terminus) resulted in a deconstructive secondary structure that was expected.
Not only does the cysteine have an effect on the secondary structure of the peptide itself, it seems to have an effect on how the peptide will interact with the ion as well.

Solvent effects have a clear effect on the secondary structure. Peptides that are hydrophilic have trouble forming secondary structures in gas phase environments without the presence of a charged species. However, these peptides easily form strong alpha helical structures in aqueous environment.
Chapter 3: Using Molecular Dynamic Simulations on the 638 Peptide to Establish a Hierarchy of Secondary Structure Hydrogen Bond Patterns

Abstract

Protein structure and protein folding is an area of extensive research. Science continues to push the area forward, but as many questions are answered, it also continues to provide questions as to why and how protein folds. Depending on the scale of protein structure, there are different forces involved in the folding and formation of each level of structure, as well as the stability for each specific type of secondary structure (alpha helix, 310 helix, beta sheets, etc.). While it is largely agreed that tertiary protein structure is the result of hydrophobic packing, it is thought that hydrogen bond patterns are the main driving force behind secondary structure. Because small peptides are not readily capable of hydrophobic packing, yet form secondary structures, there must be a driving force behind this folding. Likewise, proteins still undergo folding in environments lacking polar solvents, thus negating the hydrophobic effect. If hydrogen bonding is truly a driving force of protein secondary structure, it should be possible to determine any changes in secondary structure characteristics by altering hydrogen bond patterns. Molecular Dynamics (MD) simulations makes it very convenient to illicit secondary structural changes in peptide fragments and investigate even very subtle changes to the hydrogen bond patterns.43
Introduction

Clearly, there is a link between protein structure and function. Proteins of a similar structure will exhibit similar function and vice versa. This fact becomes a key component for any protein design. While it is not always exactly known why and how proteins fold, it is thought that the primary structure (the sequence of amino acids) is the most important component that determines the final three dimensional structure a protein forms. There seems to be something intrinsic in the amino acid sequence that induces folding events. Secondary structural features of proteins are driven by hydrogen bonding and each secondary structure feature corresponds to a specific pattern of hydrogen bonds. For example, alpha-helical structure results from hydrogen bonds forming between the (i -> i+4) residues of a protein. Similar to the alpha helix is the 310 helix which results from the (i -> i+3) residues hydrogen bonding with one another. Beta sheets also form where hydrogen bonding occurs between two residues on separate, adjacent, chains of protein. Because alpha helices make up a large percentage of secondary protein structure, and thus stabilize large sections of any protein, it is important to further understand the driving force behind this hydrogen bond formation. Through the use of Molecular Dynamic (MD) Simulations, it is possible to investigate the secondary structure of a protein and how the corresponding hydrogen bond patterns are formed.

It has been shown that the alpha-helical nature of a protein is stabilized by the concerted formation of a dipole from the component residues. Experimentally, alpha-helical peptides have been formed by placing charged sites near the C-terminus. It has
also been shown experimentally that sodiated polyalanines exhibit this effect, resulting in the formation of alpha-helical structures from globular proteins.\textsuperscript{13} Computational studies have also shown similar effects. Cysteine capping has been shown to form alpha-helical structures, from otherwise non alpha-helical sequences. It has also shown to stabilize already existing alpha-helical nature.\textsuperscript{12}

The YAP1 Protein is a biologically important protein having been linked to cancer and is thought to be a potent oncogene. It has been shown to be overexpressed during cancerous stages of a cell's life.\textsuperscript{30} YAP1 is a 65 kDa protein made consisting of 454 amino acids and 4 domains.\textsuperscript{48,49} It contains a specific domain of interest, 1SSE (residues 279-313 for Chain A, residues 565-650 for Chain B. Chain B may be a separate protein of complex?), which has 4 helices of interest.\textsuperscript{33} One helix contains no Cysteine residues present in the sequence, one helix with a Cysteine at the C-terminus of the sequence, one helix with a Cysteine at the N-terminus of the sequence, and one helix with a Cysteine residue located in the middle of the sequence.\textsuperscript{42} The fact that this protein is biologically important, and has a wide range of naturally occurring secondary structures, and thus many sequences to choose from, makes it a good model to use in order to study hydrogen bonding influences on protein folding and structure.

Using MD simulations, the native secondary structures of these helical YAP1 peptide fragments were analyzed. The native structures were compared to the structures of the same peptides under the influence of dipole manipulation. This was achieved by the placement of a charged species - a sodium ion - at the C-terminus of the
peptide. The resulting secondary structure changes, and with this change in structure, the hydrogen bond pattern is altered as well.

Methods and Materials

**Peptides of interest.** All simulations were performed using secondary structure features from the 1SSE domain of the YAP1 protein. The initial structure of the domain was retrieved from the RCSB Protein Data Bank (PDB). This structure was determined experimentally using Solution NMR and Crystallography techniques. Using the results from preliminary simulations of all alpha-helical peptides of the 1SSE domain, peptide 638 was specifically chosen for all production simulations. While peptides containing cysteine provided inconclusive results, peptide 638 provides clean results, making it a prime subject as a starting point. The lack of a cysteine allowed for fewer complications, and thus, can be used for simulations of a much longer timescale.
Figure 24: 3D Representation of the 638 Peptide Fragment
3D representation of the peptide. Residue type color coded and α-carbons for each residue labeled. (Left) Sequence view. (Right)
The initial structures of each peptide used for the MD simulations were extracted directly from the PDB coordinate files of the full protein 1SSE domain. Only the chain of interest was used to create a new PDB coordinate file of the peptide. The initial structure is used as the starting point for a series of equilibration simulations and the resulting energetically stable structure from the equilibration simulation is then used as a new starting point to perform all experimental simulations.

All PDB coordinate and PSF structure file combinations were created using VMD’s **PSFGEN** command. The equilibrated, energetically favorable structures from the minimization steps are used as the starting points for subsequent experimental simulations. After being loaded into VMD as part of a trajectory, the PSFGEN allows for the creation of clean PDB/PSF pairs to be used for further simulations.

**Computational methods.** All MD Simulations were performed in NAMD (NAnoscale Molecular Dynamics Program)\(^{35}\) using CHARMM (Chemistry at HARvard Macromolecule Mechanics)\(^{37,50}\) force fields. Analysis was accomplished through scripting/coding/programming as well as the additional use of Visual Molecular Dynamics (VMD)\(^{36}\) program. Plots were created using the Grace and Plot2 plotting programs. Additional analysis was done using Simulaid.\(^{51}\)

**Experimental Setup**

**Equilibration simulations.** All simulations were conducted in vacuum environments at 310 K. Simulations were performed at constant temperature using Langevin dynamics. Equilibration experiments were 2 ns in length, with minimization occurring for 1 ns and an additional 1 ns for equilibration. The simulation is performed
for 10,000 steps with each step representing 2 fs. The DCD - binary trajectory files - frames were collected every 1 ps with energies and pressures both collected every 0.2 ps. The restart file resulting from the equilibration simulations are used as the starting point for all subsequent experimental simulations.

**Ionization: addition of a sodium ion.** Step taken in order to add a sodium ion to the simulation environment are discussed extensively in Appendix A.

**Control/Experimental simulations.** Experimental simulations were 10 ns in length with 2 fs timesteps. DCD frames were collected every 1 ps, while energies and pressure were both collected every 0.2 ps. 15 independent experimental simulations were performed and the results then combined to obtain a numerically-averaged result.

**Results and Discussion**

**Preparation and equilibration simulations.** [Appendix B]

**Control simulations.** As a control, the fragment Peptide 638 was placed alone in a vacuum environment. The fragment is extracted directly from the PDB structure file of the entire domain. The equilibration simulation was performed at a constant temperature of 310 K and a pressure of 1 atm. After a 2 ns equilibration simulation a 10 ns MD simulation was performed, with no other forces or outside stimuli acting upon the peptide, the resulting structure and trajectory are analyzed. Using VMD, a visual inspection of the peptide was performed. A three dimensional representation of the peptides secondary structure features were observed.
Figure 25: Secondary Structure Timeline of 638 (Full Production MD)
Secondary Structure timeline of the 638 peptide only in the gas phase during full production simulation.
Figure 26: Hydrogen Bond Timeline of 638 (Full Production MD)
Hydrogen Bond timeline of the peptide only in the gas phase during full production simulation.
**Experimental simulations.** The stable structure of the fragment Peptide 638 after the control simulations was used for all experimental simulations. Experimental simulations were performed in the gas phase at 310 K and 1 atm. After proper equilibration, a sodium ion was placed at the C-terminus for a series of experimental simulations under the same conditions as the control experiment.
Figure 27: Secondary Structure Timeline of 638 and SOD (Full Production MD)
Secondary Structure timeline of the 638 peptide and the sodium ion at the C-terminus in the gas phase during full production simulation.
Figure 28: Hydrogen Bond Timeline of 638 and SOD (Full Production MD)
Hydrogen Bond timeline of the 638 peptide and the sodium ion at the C-terminus in the gas phase during full production simulation.
3-D comparison between the control and experimental simulations (Visual investigation of the ion's effect on secondary structure). Using VMD to visually investigate, and compare, the control and experimental simulations shows the main differences in structure. The extent of the alpha-helix shifts from the N-terminus toward the C-terminus. This is caused by the interaction between the sodium ion and the C-terminus of the peptide, particularly the carboxyl groups. The positive charge of the ion stabilizes the 3 partially negative exposed carboxyl oxygens at the C-terminus. This interaction forms "pseudo-hydrogen bonding", directly resulting in a shift in the alpha-helical structure.
Figure 29: 3-D Comparison of the Sodium’s Effect on Peptide 638
3-D representations of the peptide secondary structure with (top left/right) and without the ion (bottom left/right)
A helical peptide exhibits a dipole due to the coordination of the individual amino acid dipoles, from the partially positive amine to the partially negative carboxyl groups. This helical macro-dipole thus interacts with the positive charged ion, creating a partial charge/dipole interaction.
Table 2: Active Hydrogen Bonds in the 638 Peptide MD Simulations
A comparison of active hydrogen bonds from the 638 peptide with and without the ion present.

<table>
<thead>
<tr>
<th>Hydrogen Bond</th>
<th>Carboxyl Oxygen Residue</th>
<th>Amine Hydrogen Residue</th>
<th>Atom Numbers (N-H---O)</th>
<th>Present in the Control Simulations (Above 50%)</th>
<th>Present in the Experimental Simulations (Above 50%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB1</td>
<td>638</td>
<td>642</td>
<td>73-74---29</td>
<td>= ~50</td>
<td>No (0%, not present)</td>
</tr>
<tr>
<td>HB2</td>
<td>639</td>
<td>643</td>
<td>90-91---44</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>HB3</td>
<td>640</td>
<td>644</td>
<td>109-110---56</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>HB4</td>
<td>641</td>
<td>645</td>
<td>119-120---72</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HB5</td>
<td>642</td>
<td>646</td>
<td>138-139---89</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HB6</td>
<td>643</td>
<td>647</td>
<td>152-153---108</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HB7</td>
<td>644</td>
<td>648</td>
<td>174-175---118</td>
<td>= ~50</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Hydrogen Bond Analysis

**Graphical analysis.** Two main features were investigated using VMD's Timeline feature: a secondary structure timeline and a hydrogen bond timeline are created. First, a trajectory from the NAMD simulations was loaded into VMD and the timeline produces a graphical representation of the secondary structure at every time step of the trajectory. The graph plots the secondary structure that occurs between each residue in the primary sequence (the y-axis) at each individual time step (the x-axis). When each step is processed, the resulting information is compressed, producing a single graphical representation of the changing secondary structure features of the peptide throughout the length of the simulation. VMD uses a color-coding system to distinguish each type of secondary structure feature. A green representation indicates a coil or simple turn, blue indicates a 310 helix, and a pink representation indicates the presence of alpha-helical structure. Any white space indicates no discernible secondary structure features. In addition, red represents a Pi helix, while yellow indicates a Sulfur-Sulfur bridge forming between Cysteine residues. However, in the peptide of interest, no Pi helix or disulfide bridges were present.

Similar to secondary structure timelines, hydrogen bond timelines can give complimentary information. The resulting graphs indicate the presence or absence of a hydrogen bond. As with the secondary structure timeline, the x-axis remains the same (the frame number of the trajectory), while the y-axis shifts from each residue in the primary sequence to a list of three atom codes (heavy-heavy-small) involved in the hydrogen bond. Each number corresponds to an atom that makes up that specific
hydrogen bond. The first number is the nitrogen of the amino acid amine (N-H). The second number is the oxygen of the amino acid carboxyl group (C=O). The third number is the hydrogen of the amino acid amine (N-H). Together, these are the three atoms that form a hydrogen bond between the backbone atoms of the residues (C=O----H-N).

The presence of a hydrogen bond is also color-coded using VMD's coding system. All areas which are black indicate no hydrogen bond present at a given point - that particular hydrogen bond was not present, or "off", at that particular frame. All areas which are white, indicate that the hydrogen bond were present, or "on" at that particular frame. Quantitative data for each hydrogen bond timeline can be obtained in a list; the graph uses numerical data for the "on/off" representation. Using binary data, a 1 or 0 is used to indicate whether the individual hydrogen bonds are on or off, respectively.

**Hydrogen bond timeline parameters.** In order to use VMD's hydrogen bond timeline to better indicate the hydrogen bonds of interest, parameters must be set appropriately. Because backbone hydrogen bonding is the primary area of interest, other hydrogen bonds must be first filtered out, thus displaying only the hydrogen bonds that are believed contribute to the secondary structure of the peptide. In order to accomplish this, only hydrogen bonds between N (the amine nitrogen), HN (the amine hydrogen), O (the carboxyl oxygen), and C (the carboxyl carbon) are displayed. The hydrogen bond distance is set to 3.5 angstroms and the angle is set to 45 degrees, as per the VMD User guide.52,53
Control. Analysis of the control simulation’s secondary structure timeline indicates two key pieces of information; there is no stable alpha-helical nature present between residues 644 and 648, and the alpha-helical nature between residues 637 and 644 is not robust. There is some 310 helical nature between these residues throughout the simulation, along with the alpha-helical nature, and even some coils. These secondary structure features alternate throughout the simulation. The hydrogen bond timeline of the control simulation indicates a comparable pattern. The hydrogen bonding is transient, and it should also be noted that five hydrogen bonds seem to be primarily responsible for the majority of the contribution to the secondary structure features.

Experimental simulation. The addition of a sodium ion near the C-terminus of peptide 638 resulted in two key, and very noticeable, differences from the control simulations of the peptide without the ion nearby. First, the alpha-helical structure undergoes a location shift. The secondary structure timeline indicates a strong stabilization of the alpha-helical nature of the peptide toward the C-terminus. From residues 641-647, the alpha-helical nature is increased, and preserved throughout the entirety of the simulation. However, the alpha-helical nature that is present in the control experiments from residues 638-640 is completely lost. The resulting structure from residues 638-640 is mostly turn/coil. Although the alpha-helical nature shifts, it can still be inferred that the overall alpha-helical nature is stabilized in peptide 638, albeit, toward the C-terminus. This can be thought of in the sense that the alpha helix spans between 7 residues in the experimental simulations with the ion (residues 641-
as opposed to the 6 residues in the control experiments (638-643). Qualitatively, the secondary structure timelines for the alpha-helical nature of the experimental simulations indicates more stability as well. The timeline shows more consistent alpha-helicity throughout the simulation, with few occurrences of different secondary structure features. In addition to the secondary structure timelines, the hydrogen bond timelines also reflect this change. The number of prominent hydrogen bonds increases from the three or four in the control simulations, to four or five in the experimental simulations. The hydrogen bond timeline continues this trend in stability as the prominent hydrogen bonds remain on more consistently throughout the simulation. It can be inferred that the alpha-helical nature, and the hydrogen bond pattern, shifts toward the C-terminus as a direct result of the ion. As the helix now forms on the same end of the peptide where the sodium ion was placed.

Comparing structural features of the control/experimental simulations (quantifying secondary structure). Using scripting to data process the secondary structure data files, it is possible to quantify the alpha helical nature of the peptide. The peptide consists of 12 residues and the simulation lasts for 100,000 total frames. This equates to a total of 120,000 total possible secondary structure features of the peptide throughout the length of the simulation. Each of the 15 simulations for the two types (without the ion and with the ion) is processed to count each time alpha-helical nature occurs.
Table 3: Quantifying Secondary Structure Properties of the 638 Peptide
Total counts of alpha helical secondary structure features throughout the simulation.

<table>
<thead>
<tr>
<th>Simulation Number</th>
<th>Counts of alpha helical nature</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without the Ion Present</td>
<td>With the Ion Present</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>69,060</td>
<td>68,181</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>67,144</td>
<td>69,136</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>56,638</td>
<td>67,864</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>69,602</td>
<td>68,203</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>63,895</td>
<td>69,433</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>66,711</td>
<td>68,325</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>72,661</td>
<td>68,926</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>66,820</td>
<td>68,658</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>65,540</td>
<td>68,227</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>60,495</td>
<td>70,369</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>40,825</td>
<td>68,266</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>56,874</td>
<td>68,531</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>65,235</td>
<td>68,222</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>73,716</td>
<td>68,536</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>69,019</td>
<td>69,346</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>64,282</td>
<td>68,682</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>8,167.2</td>
<td>656.28</td>
<td></td>
</tr>
<tr>
<td>RSD</td>
<td>12.71%</td>
<td>0.96%</td>
<td></td>
</tr>
<tr>
<td>Total Percent Alpha-Helicity</td>
<td>53.568%</td>
<td>57.235%</td>
<td></td>
</tr>
</tbody>
</table>
The quantification of alpha-helicity results in two conclusions. First, the increase in total alpha-helicity from 53.6% to 57.2% shows the stabilizing effect of the ion on the alpha-helical nature of the peptide. However, even more convincing is the drastic decrease in the relative standard deviation, which falls from 12.71% to 0.96%. This is a change of about 3.5% for absolute and about 6% for relative. This quantitatively shows a greater stabilization in the overall alpha-helical structure by the sodium ion.

**Negative control.** In order to determine that the alpha-helical stabilization is in fact due to the ion placement, and not just the presence of the ion, the sodium ion is also placed near the N-terminus of the control peptide. The simulation parameters of the experimental simulations are retained, with the only difference being the placement of the ion. The resulting secondary structure is noticeably altered, producing an extremely unordered peptide. As opposed to the stabilizing effect at the C-terminus, the alpha-helical nature was destroyed, indicating the placement of the ion at the N-terminus to be more destructive when compared to the constructive nature when placed at the C-terminus. This could be expected as the ion in this position opposes any helical macro dipole that could be formed.
Figure 30: Secondary Structure/Hydrogen Bond Timelines for Negative Control

Secondary Structure (top) and Hydrogen Bond (bottom) timelines for negative experiments.
Removal of the sodium ion after helical stabilization of the peptide. It is clear that the sodium ion has a stabilizing effect on the alpha helical nature of the peptide when placed near the C-terminus. With the change in secondary structure, a shift in hydrogen bond pattern also occurs. However, the exact effect of the ion is questioned. So the ability of the peptide to retain its shifted alpha-helical secondary structure without the ion is also tested. The energetically stable structure of the peptide and the ion is still used as the starting point, however the ion is removed. The resulting lone peptide is then used as the starting structure for a new set of experimental simulations. What occurs is an alpha helical secondary structure near the C-terminus that is much more relatable to the secondary structure of the peptide with the ion than the secondary structure of the peptide without the ion. However, it is not fully retained, as it seems the stability of the structure is not as sound. This change is also reflected in the hydrogen bond pattern.
Figure 31: Timelines after the Removal of the Sodium Ion (Helix Stabilized)

Timelines for the 638 Peptide with removal of the sodium ion after helical stabilization has been established
Table 4: Quantifying Alpha-Helical Nature After the Removal of the Sodium Ion

Alpha helical nature of the peptide after stabilization/removal of the sodium ion.

<table>
<thead>
<tr>
<th>Simulation Number</th>
<th>Counts of alpha helical nature (Removed ion from stabilized peptide)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65,988</td>
</tr>
<tr>
<td>2</td>
<td>65,743</td>
</tr>
<tr>
<td>3</td>
<td>66,221</td>
</tr>
<tr>
<td>4</td>
<td>66,630</td>
</tr>
<tr>
<td>5</td>
<td>65,379</td>
</tr>
<tr>
<td>6</td>
<td>69,169</td>
</tr>
<tr>
<td>7</td>
<td>68,658</td>
</tr>
<tr>
<td>8</td>
<td>65,589</td>
</tr>
<tr>
<td>9</td>
<td>66,527</td>
</tr>
<tr>
<td>10</td>
<td>69,070</td>
</tr>
<tr>
<td>11</td>
<td>66,623</td>
</tr>
<tr>
<td>12</td>
<td>66,800</td>
</tr>
<tr>
<td>13</td>
<td>65,447</td>
</tr>
<tr>
<td>14</td>
<td>66,622</td>
</tr>
<tr>
<td>15</td>
<td>66,637</td>
</tr>
<tr>
<td>Average</td>
<td>66,740</td>
</tr>
<tr>
<td>SD</td>
<td>1,247.4</td>
</tr>
<tr>
<td>RSD</td>
<td>1.87%</td>
</tr>
<tr>
<td>Percent Alpha-Helicity</td>
<td>55.617%</td>
</tr>
</tbody>
</table>
This indicates that while the ion does have a great stabilizing effect on the peptide and does cause a shift in structure (and therefore in hydrogen bond pattern as well) it is not necessary to continue to stabilize that structure. Once the structure - and hydrogen bond pattern - is in place that is sufficient to retain such a secondary structure.

**Quantitative hydrogen bond data.** While the qualitative nature of the structural changes is intriguing, it is also possible to use a quantitative approach to explain these secondary structural changes. In order to achieve better even statistics, 15 independent simulations were performed and averaged for both the control and experimental simulations. Secondary structure timelines and hydrogen bond timelines were produced for every simulation. Using the hydrogen bond timelines, it is possible to determine and choose which hydrogen bonds are more prominent, and most likely to contribute to the secondary structure features of the peptide. These specific "main" hydrogen bonds were chosen and isolated. Because of the binary fashion of the data output for the hydrogen bond timelines, it is possible to average the lifetime of the specific hydrogen bonds for all 15 simulations. The individual hydrogen bond receives a 1 or a 0 (on or off respectively) for every frame. This data is extracted from the output file for each hydrogen bond. The value at every frame, for each of the 15 simulations is averaged, producing a value between 1 and 0 for each frame. When the average for every frame is combined (in similar fashion to a timeline) a quantitative representation of the life for each main hydrogen bond can be obtained.
Figure 32: Quantifying Lifespans of Individual Hydrogen Bonds
Tracking the life of individual hydrogen bonds through the length of a simulation
Once again, two key pieces of information are obtained. The simulation of the peptide with the ion stabilizes the alpha helix, and with it, the four main hydrogen bonds. The four isolated hydrogen bonds are dramatically stabilized. There is much more stability in these hydrogen bonds when the ion is present, as opposed to when there is no ion present. Two points can be inferred from this. First, the fact that the hydrogen bond is much closer to “on” (a value of 1) from the start of the simulation and throughout the length of the simulation is clear. Second, the difference in noise between the two is also apparent.

The second piece of information gained from this tracking technique, the belief that the structure/hydrogen bonding of the stabilized peptide after the removal of the sodium ion was somewhere in between the control and experimental simulations is also confirmed. With each of the four prominent hydrogen bonds, in every case, the simulations of the removed ion land in the middle of the control and experimental simulations.

One piece of information noticed is that there is a difference among common hydrogen bonds, i.e. some form better with the ion and some form better without the ion. However, it seems that one hydrogen bond in particular forms without the ion, but does not for at all when the ion is present. This may be a hydrogen bond to isolate for future experiments.

This allows the ability to choose specific hydrogen bonds to be determined responsible for specific secondary structure features. This can be used later for MD simulations in an attempt to fold a peptide. This can be used to not only fold a peptide
back to a native structure, but also a desired structure, since the peptide has two different structures between the simulations with the sodium ion and without the sodium ion.

**Conclusion**

Using a series of MD simulations, it is possible to investigate secondary structural changes in small peptide fragments and determine any changes in hydrogen bond patterning that may result. Using MD simulations, the native structure of the YAP1 peptide fragment, peptide 638, was determined in a vacuum environment. It was shown that the alpha helical nature of the peptide resides toward the N-terminus. It was also revealed that the total secondary structure was highly variable through the series of repeated experiments. By adding a charged species near the C-terminus, it was possible to shift the alpha helical nature toward the C-terminus. Not only causing a shift in structure, but causing an increase in stability. The changes in structure and stability were then explained through the investigation of the hydrogen bond patterns between the two differing structure.
Chapter 4: Steered Molecular Dynamic Simulations of the 638 Peptide Fragment to Illicit Unfolding Events in Order to Attempt to Better Understand the Folding Pathway

Abstract

Protein folding is a complex process that involves many factors and many different forces. Depending on the level of protein structure being studied, the main driving forces behind protein folding can differ. Generally, it is believed that the process of hydrophobic packing mostly drives the folding of proteins at the tertiary structure level forward. While this provides a very reasonable explanation for the folding and formation of "larger" protein structures, it is still possible for peptides to form three-dimensional structures. In addition to this, it is also possible for proteins, and peptides, to form three-dimensional structures in the gas phase, eliminating any solvent packing effects. Because of this, hydrogen bonding and hydrogen bond patterns are suggested as the main driving force behind secondary structure.

Introduction

Because proteins fold by levels, it is necessary for the secondary structures to form before the larger protein forms a tertiary structure. As tertiary structure is the result of interacting secondary structures, secondary structures will have an effect on the final three-dimensional shape of a protein, making it very important to further explore the driving force of secondary structure. While there are few experimental
methods able to explore the dynamics behind protein folding on such a small scale, there are many theoretical techniques that can help attribute explanations for protein folding. Particularly, Steered Molecular Dynamics (SMD) simulations can provide an in-depth look at the applied forces to biological macromolecules. Because SMD simulations can provide better perspective at the forward, unfolding pathways of peptides, it can be used to investigate the reverse, or folding pathways of peptides.

While progress is constantly being made into the area of why and how proteins fold, it is still not 100% clear. If a desired structure is needed, a sequence cannot be generated from scratch to provide this structure. Likewise a random sequence from a resulting structure cannot be predicted with certainty. While the overall picture is still unclear, there are several driving forces in protein structure that are thought of as the main culprits of protein folding. For secondary protein structure, it is possible that hydrogen bonding and hydrogen bond patterns are the main driving force. Each secondary structure feature has a specific hydrogen bond pattern. For tertiary structures, the main driving force is hydrophobic/hydrophilic interactions. Because proteins are made up of amino acids that can be polar or non-polar in nature, these residues will behave differently in response to their environment. In aqueous environments, polar groups will want to be in areas of polar solvent (likes dissolve likes) (hydrophilic) and non-polar groups will want to be away from any areas of polar solvent (hydrophobic). This dichotomy results in proteins folding in such a way, that polar groups are exposed to polar solvent, while non-polar groups are hidden away.
Although tertiary structure is the result of hydrophobic collapse, small peptides are able to exhibit and form secondary structure features. These smaller peptides cannot easily undergo folding events to pack away hydrophobic regions due to size limitations. There must be a force that applies to secondary structure aside from hydrophobic collapse. If hydrogen bond formation is truly a driving force behind secondary structure, it should be possible to fold small peptide fragments through the preservation of specific hydrogen bonds known to contribute to specific secondary structure features of a peptide. Steered molecular dynamic (SMD) simulations allow the user to apply forces to simulations of biochemical systems. These forces can be applied to specific atoms, allowing these atoms to be "pinned" into place. Or a directional force can be applied, allowing the atom to be pulled, or pushed, in a specified direction. Through the use of SMD simulations, peptides of interest will be pulled in such a way that the secondary structure is completely destroyed. This information will then be used in an attempt to fold these same peptides to their natural structures or a desired non-native structure, and a hydrogen bond hierarchy may be established. Manipulating these hydrogen bonds may be possible in order to stimulate folding events using standard molecular dynamic simulations. Because these pulling simulations can provide information about unfolding events, the information obtained may be used to determine the forces in the opposite direction, i.e., folding events.

**Methods and Materials**

**Peptides of interest.** The peptide fragment 638 from the 1SSE domain of the Yes-Associated Protein (YAP1) Protein was the peptide of interest used during all
simulations performed. The 638 peptide was previously used during Molecular Dynamic (MD) simulations. The MD simulations provided structural characteristics and hydrogen bond patterns for the 638 peptide in gas phase environments. The comparison of the altered secondary structure features, achieved by the placement of charged species near the peptide, determine changes in the secondary structures as well as the resulting change in the hydrogen bond pattern (Chapter 3). The molecular dynamic simulations provide familiarity with the native structure and hydrogen bond patterns which can be used in tandem with the SMD results to better predict which structural features and which hydrogen bond patterns can be used in attempts to fold the peptide of interest. A hydrogen bond hierarchy has already been established for the 638 peptide in Chapter 3 and the initial structures will be used as the starting point of the SMD simulation.

Computational methods: NAMD/VMD programs & SMD simulations. NAMD was used for the molecular dynamic simulations, while VMD is used for visual investigations of the resulting trajectories. All SMD simulations were performed under constant velocity with two separate types of constant velocity simulations being employed. TCL constant velocity simulations were used for additional output files, while configuration constant velocity simulations were employed for primary pulling simulations. The spring constant for these simulations is set to 7.2 (pN/A) and a velocity of 0.00001 A/timestep is utilized.

Experimental Setup

Equilibrations experiments. In order to perform SMD simulations, equilibration MD experiments must first be carried out. Details for these steps are located in
Appendix A. Using Gas-Phase MD simulations, the peptide and the ion are equilibrated. The ion is placed near the C-Terminus; the equilibration allows the ion to become situated into a pocket formed by the free C-Terminus oxygen from the three carboxyl groups. This equilibration allows a favorable ion-dipole interaction with the helix macrodipole, stabilizing the alpha-helical structure. If the SMD simulation is performed without equilibration, i.e., a direct pull on the ion prior to stabilization, the peptide is pulled and the breaking pattern of the helix varies, providing irreconcilable results, i.e. there are no smooth transition breaks of the alpha helix. This once again shows the stabilizing effect of the ion on the peptide because the secondary structure features are being destroyed far too quickly.
Figure 33: Secondary Structure Timeline for SMD Simulation Without Equilibration

The Secondary Structure Timeline of the SMD simulation of Peptide 638 and sodium ion without equilibration. The Secondary Structure is destroyed with no noticeable trend.
Alignment. For simplicity of applied forces and direction, the peptide and ion are aligned with the z-axis: The system is placed in such a way that the Z-axis runs through the center of the helix. The C-terminus of the peptide, as well as the sodium ion, are placed in the positive z-direction. While the N-terminus is placed toward the negative z-direction, or the origin.

Applied forces. The α-carbon at the N-terminus is fixed in space in order to keep it from moving. The location of this anchor point allows for entirety of the peptide to be involved in the dynamics of the simulation. The pulling force is applied only on the sodium ion in the (mostly) positive z-direction. The precise direction of the force is determined exactly using vectors (TCL/SMD Simulations).

Results and Discussion

"Control simulations". Previously, the secondary structural features of the 638 peptide have been determined with a sodium ion at the C-terminus of the peptide and without a sodium ion. In both cases, the peptide exhibited an alpha-helical structure. Without the ion in the system, the alpha helix was retained near the N-terminus of the peptide and was not clearly stable along the whole length of the peptide. However, once the sodium ion was added to the system, the alpha helix shifted toward the C-terminus, and become visibly more stable. Further investigation showed that with the change in secondary structure, a shift in the hydrogen bond pattern also occurred. These two key results from the MD experimentation of the 638 peptide will be used in conjunction with SMD simulations to further investigate folding events in the 638 peptide.
Figure 34: Secondary Structure Timeline for Peptide 638 Without Sodium Ion
The secondary structure timeline of peptide 638 from the production MD simulations. Previously determined. There is a fair amount of alpha helical nature is present toward the N-terminus of the peptide.
Figure 35: Secondary Structure Timeline for Peptide 638 With Sodium Ion
The secondary structure of the peptide and the sodium ion at the C-terminus from the MD production simulations. The alpha helical nature shifts to the C-terminus, the area with the sodium ion present.
Folding Extended Peptides Using Standard MD

**Molecular dynamics simulations to fold peptides.** While it is possible to fold a protein using standard molecular dynamic simulations, it can require a large amount of computing power\(^6\) or a large amount of computing time. Novel approaches can also be used to fold proteins (replica exchange\(^6\), fast folding proteins used\(^7\), pressure jump\(^8\), force field bias\(^9\)). These simulations will reduce the amount of time/resources needed, but come with "synthetic" drawbacks.\(^\text{10}\)

In order to demonstrate this point, a folding simulation was attempted. The 638 peptide is placed in a gas-phase environment and fully extended using steered molecular dynamics (SMD) simulations. This fully extended peptide is then used as the starting point for a standard gas phase molecular dynamics simulation. The extended peptide undergoes a 2 ns equilibration followed by a 10 ns MD simulation (as well as 100 ns simulation). The results show that this is not enough to fold the peptide into its native structure. While the peptide does fold, it is very disordered, and does not resemble the alpha helical native structure that the gas phase MD simulations of the 638 peptide alone shows.
Figure 36: Secondary Structure Timeline for Peptide 638 in an Attempted Folding

Secondary Structure Timeline of the fully extended 638 peptide (top). This was the secondary structure timeline for the attempted folding MD equilibration simulation. 3D structures of the peptide at the start (bottom – left) and end (bottom – right) of the simulations.
The hydrogen bonds expected to be responsible for the alpha helical nature are investigated using similar methods to previous hydrogen bond analysis. What is observed is that many hydrogen bonds thought to be responsible for the alpha helical nature do not reform during the folding simulations using standard molecular dynamics methods. However, even the few hydrogen bonds that do reform do not do so with high stability, as seen in the low percentage of the lifespan.
Figure 37: Hydrogen Bond Lifespans During Attempted Folding Simulations
The lifetime of isolated hydrogen bonds during the folding of the fully extended 638 peptide while using standard molecular dynamic simulations.
Experimental Simulations:

Snap-Back re-fold SMD simulations (SMD simulations of peptide with ion).

After the series of MD simulations to prepare the necessary environments, as well as the MD simulations used to test the folding capabilities of the 638 peptide, a series of SMD simulations are performed to observe unfolding events of the peptide. The \( \alpha \)-carbon of the N-terminus of the peptide has a force applied to it, which constrains it to its location. Another force is applied to the sodium ion in the positive z-direction. Because the peptide is oriented with the N-terminus at the negative z-direction and the C-terminus/ion at the positive z-direction, the force applied to the ion results in a pull of the C-terminus/ion away from the N-terminus in a (mostly) straight line/vector. The pull on the sodium ion causes a charge/partial charge interaction of charged sodium ion and C-terminus of peptide (partial negative oxygen). This interaction results in the peptide being pulled apart as the ion is pulled away from the peptide in the positive z-direction. At a point in the SMD simulation, the peptide will become fully extended, resulting in a removal of all backbone hydrogen bonding. As the peptide is fully extended, the charge/partial charge interaction between the sodium ion and C-terminus of the peptide will become more difficult to maintain. Eventually, this interaction between the ion and peptide breaks, the peptide “snaps-back” on itself, and undergoes a folding, or "re-folding" event. The ion continues to be pulled away and the simulation is allowed to finish. Analysis of the trajectory of the SMD simulation yields very interesting results. The initial secondary structure of the SMD trajectory (roughly the first half of the simulation/the point before full extension of the peptide) is very reminiscent of the
secondary structure for the peptide + ion system during the MD simulations. However, the secondary structure of the peptide after the re-folding event (second half of the SMD simulation/after the interaction between the sodium ion and the C-terminus breaks/ceases) is much more reminiscent of the secondary structure of the peptide during the MD simulations WITH OUT the ion MD simulations. This reproduces the stabilization effect of the ion on the alpha helical nature of the peptide. In addition, the results of the SMD simulation exhibit a possible folding pathway for the 638 peptide.
Figure 38: Hydrogen Bond and Secondary Structure Timeline for Peptide 638 with Sodium SMD Simulations

Secondary Structure/Hydrogen Bond timelines for SMD/Re-fold simulations. The peptide is pulled beyond full extension and folds back on itself. The Secondary Structure before full extension and collapse is noticeably different.
Figure 39: Comparing the Hydrogen Bond Lifespans Between Successful and Failed Re-folding simulations

Comparison of the hydrogen bonds of interest during SMD simulations. The reforming of specific hydrogen bonds is investigated in successful re-folding simulations (Red-line) and unsuccessful re-folding simulations (Blue-line).
Using SMD simulations to re-fold Peptide 638 yields varying results. In some cases, the peptide is fully extended, the interaction between the ion and C-terminus of the peptide breaks, and the peptide re-folds to an alpha helical nature. However, other times, the alpha helical nature does not reform after the snap-back. Comparing the reformation of hydrogen bonds between the successful and unsuccessful refolding techniques allows a better understanding of the hydrogen bonds responsible for the alpha helical nature.
Table 5: Active Hydrogen Bonds in the 638 Peptide During SMD Simulations
Active hydrogen bonds from the 638 peptide during SMD Simulations (Before and After full extension/snap-back folding)

<table>
<thead>
<tr>
<th>Hydrogen Bond</th>
<th>Carboxyl Oxygen Residue</th>
<th>Amine Hydrogen Residue</th>
<th>Atom Numbers</th>
<th>Present During SMD simulation prior to extension (\textasciitilde with ion)</th>
<th>Present during SMD simulation after full extension (\textasciitilde without ion)</th>
<th>Present in the MD simulations with the ion</th>
<th>Present in the MD simulations without the ion</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB1</td>
<td>637</td>
<td>641</td>
<td>Start</td>
<td>No</td>
<td>~50%</td>
<td>No</td>
<td>~50%</td>
</tr>
<tr>
<td>HB2</td>
<td>638</td>
<td>642</td>
<td>73-74---29</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HB3</td>
<td>639</td>
<td>643</td>
<td>90-91---44</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HB4</td>
<td>640</td>
<td>644</td>
<td>109-110---56</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HB5</td>
<td>643</td>
<td>647</td>
<td>152-153---108</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HB6</td>
<td>644</td>
<td>648</td>
<td>182-183---118</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>HB7</td>
<td>641</td>
<td>645</td>
<td>119-120---172</td>
<td>~50%</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>HB8 (3₁₀ helix)</td>
<td>641</td>
<td>644</td>
<td>109-110---56</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
One question arises; if secondary structure and hydrogen bond patterns are directly linked, do the hydrogen bond patterns of the first half of the SMD simulation match up with the hydrogen bond patterns of the peptide with the ion MD simulations? Likewise, do the hydrogen bond patterns of the second half of the simulation correspond to the peptide MD simulations without the ion?

**Pin SOD ion, pull on Ca of N-Terminus.** In order to show that the order in which hydrogen bonds break is not purely a consequence of the pulling set up, a reverse pull is performed. The system is mirrored in orientation in order to place the N-terminus, which was previously toward the negative z-direction, in the positive z-direction. The sodium ion and the C-terminus of the peptide, which was previously in the positive z-direction, is now placed at the negative z-direction. All other orientations are attempted to be preserved as possible, i.e., the helical peptide is still oriented around the direct z-axis. This allows for the pulling force to be calculated, and applied, in mostly the z-direction with very little x or y directional pulling.

As expected, HB 1 and 4 break prior to 2 and 3, as seen in the SMD simulations where the N-terminus is held in place while the sodium ion is pulled. These SMD simulations provide enough information to determine that the hydrogen bonds break in an order not determined by the actual pulling. This can be attributed to a difference in strength amongst specific hydrogen bonds OR the presence of additive strengths amongst the hydrogen bonds as the two hydrogen bonds in the middle of the peptide seem to hold longer than the two hydrogen bonds on the "outside" of the peptide.
Figure 40: Reverse Pulling Simulations to Compare Hydrogen Bond Breaking Patterns

SMD simulations where the sodium ion is pulled while the N-terminus is pinned (Top).
SMD simulations where the N-terminus is pulled while the sodium ion is pinned (Bottom).
Determining Important Hydrogen Bonds

In addition to the individual SMD simulations contributing to the information about which specific hydrogen bonds are considered important, the replicates of the 15 simulations performed also reveal striking data. The 15 replicates are performed for all simulations, this is to provide better statistical coverage and sampling for the groups, but the individual results of each of the replicates can also be examined.

Because the DSS/HB timelines share a common variable - the x-axis is equal to the timestep - it is possible to overlay the two graphs in order to better understand and corroborate the connection between the hydrogen bond patterns and secondary structural features. While the other variable - the y-axis is the secondary structure type at a specific residue for the secondary structure timeline and the presence of a hydrogen bond occurring between the 3 specified atoms for the hydrogen bond timeline - are not exactly equal, they share a common connectivity since the hydrogen bond pattern and the secondary structure types are directly connected.

The 15 replicate simulations are loaded back-to-back and the two separate timeline are obtained. The two timelines are overlain and a transparency is added to the graph to supply a better view of both plots equally. Using this technique, it becomes clear which main hydrogen bonds are the most important and directly responsible for which secondary structure attributes of the peptide. This coupled along with the hydrogen bond hierarchy information previously established from the SMD simulations (pulling simulations passed full extension -> re-fold) can be used together to decide
which are the most important hydrogen bonds and which specific hydrogen bonds can be possibly preserved in an attempt to re-fold the peptide under standard MD practices.
Figure 41: Overlay of the Secondary Structure Timeline and Hydrogen Bond Timeline for the 15 Replicate SMD Simulations
Overlay of the secondary structure timeline and the hydrogen bond timeline. A transparency is added to view both timelines. Three hydrogen bond timelines show as the most responsible for the alpha helical nature of the peptide on the secondary structure timeline.
Using the information obtained by stopping the simulation just short of the interaction between the ion and peptide breaking, it can be seen which 3 hydrogen bonds are very specific to forming alpha helical features in the peptides secondary structure.

**Using SMD simulations to stimulate folding.** The peptide is pulled to full extension, as a direct result of the pulling, the hydrogen bonds amongst the backbone of the peptide also break. Using an overlay of the secondary structure timeline and hydrogen bond timeline, it is possible to see which specific hydrogen bonds are responsible for which secondary structure features. As before, the main goal is to attach specific hydrogen bonds to specific secondary structure features and use this information in an attempt to fold a peptide using standard MD simulations.

It is also possible to determine the order in which these hydrogen bonds break. The two hydrogen bonds near the termini break first, while the two hydrogen bonds in the middle of the peptide break last. Whether this is because of the strength of the individual hydrogen bonds, the combined strength of the two hydrogen bonds in the middle contributing to each other, or just a consequence of the pulling technique is unsure. This can possibly display an indication of strength/importance between the individual hydrogen bonds or can provide evidence that it is merely a consequence of pulling.

**Attempting to fold peptides by preserving hydrogen bonds.** Using the information gained from the SMD simulations, as well as the previous MD simulations, it will be attempted to fold a peptide by preserving backbone hydrogen bonds. Reference file SMD simulations are used to pull a peptide apart around the hydrogen bonds that
are desired to be preserved. The aligned peptide and sodium ion system is used as the structure file. The hydrogen of the amine (N-H) is changed to a beta value of 1 as well as the oxygen of the carboxyl group (O=C). The α-carbon of the C-terminus is set to an occupancy value of 1. This allows the peptide to be apart from the C-terminus until the oxygen of the anchored group. The system is then "flipped" around and the same pinned atoms are still held in place. But instead of the pulling force being applied to the C-terminal α-carbon, it is applied to the N-terminal α-carbon. An addition pinning force is applied to the C-terminus α-carbon in order to prevent the previously extended end of the peptide from re-folding. The resulting structure is a peptide that is fully extended on both sides of the chosen hydrogen bond(s). This extended peptide with specifically preserved hydrogen bond(s) is used as the starting structure for a series of MD simulations in an attempt to fold the peptide to a secondary structure that is more resemblant to the secondary structure formed by the peptide in a gas phase environment.
Figure 42: Secondary Structure Timeline for Folding MD Simulations (Using Preserved Hydrogen Bonds)

MD Folding Simulations when specific hydrogen bonds are preserved. One HB is preserved (top), but does not fold back to the expected secondary structure. Two HB are preserved (bottom), while only one HB is held in place. What is observed is the alpha helical secondary structure returns.
Figure 43: Hydrogen Bond Lifespan for Hydrogen Bond Folding Simulations
The lifetimes of the investigated hydrogen bonds. HB3 is held constant throughout the simulation in hopes of stimulating folding events. The majority of hydrogen bonds reform during the folding of the peptide to a more alpha helical structure, however, HB7 in particular does not reform well.
Quantitative analysis of the hydrogen bonds. Similar to the lifetime of specific hydrogen bonds analysis in the MD paper, the same tracking of individual hydrogen bonds can be performed for the SMD simulations, with similar results/interpretations. The order in which they break can be inferred (graph drops below 0.5 first, breaks first. Graph that remains above 0.5 longest, breaks last). This type of analysis can also lead to the conclusion of whether the hydrogen bond patterns of the MD simulations corresponds to the hydrogen bond patterns of the SMD simulations.
Conclusion

From previous MD simulation results, it was observed that the secondary structure of peptide 638 was alpha helical at the N-terminus of the peptide. The addition of a sodium ion stabilized the alpha helical nature at the C-terminus of the peptide. Using SMD simulations the helix can be destroyed by pulling the ion away from the peptide, the charge/partial charge interaction between the peptide and the ion causes the peptide to unfold, before eventually the interaction breaks. If the simulation is ended at full extension, it becomes obvious which hydrogen bonds are responsible for the alpha helical nature. Additionally, if the simulation is allowed to proceed past full extension, and past the broken interaction between ion and peptide, the peptide will undergo a "snap-back" process, stimulating a re-folding effect. The information gained from MD simulations can be couple with the information gained from the various SMD simulations, and a hydrogen bond importance hierarchy can be established. This hierarchy was then used to fold a fully extended 638 peptide back to a more natural helical structure, which cannot be accomplished readily using standard MD practices and a completely extended 638 peptide (completely extended to full secondary structure destruction/no hydrogen bonds remaining).
Chapter 5: An Attempt to Perform Peptide-Ion Interaction/SMD Simulations in the Aqueous Phase

Introduction

Peptides are capable of different properties dependent on the type of environment they are contained it. It has been shown that acidity of peptides can differ in the gas phase than that of the acidity of the same peptide in the aqueous phase. This difference in chemical properties is not the only difference seen in differing environments as structural changes are also observed.

While it is very important to understand how proteins fold intrinsically, which gas phase simulations help provide information towards, it is also equal important to understand the workings of protein folding in more naturally occurring environments. As biological activities occur in the body, aqueous environments, or solvents, will always be present. Hence, it is important to perform computational studies in aqueous phase environments as well, in addition the all gas phase simulations.

Methods & Materials

Parameters. Every simulation is 500,000 steps in length or 1000ps (1ns). A 500 step minimization is utilized for the aqueous phase simulations as these require more equilibration. All simulations are constant pressure, constant temperature, and constant volume. Periodic boundaries are established on the water-box based on the vectors of the sides of the water box.
**Preparation.** In contrast to gas phase simulations, the preparation of aqueous phase environments is more sensitive. The clean pdb/psf combination of the peptide of interest must first be solvated. This is typically accomplished using VMD’s solvation package. While there are additional ways to solvate, this is often the cleanest/simplest way of accomplishing solvation. After the peptide is solvated, the system is ionized with a single sodium ion. The water molecules are then removed by saving just the peptide and the sodium ion to a new psf/pdb file pair. The ion is then moved to the desired location, in most cases, near the C-terminus of the peptide. Moving the ion after removing the water works around any placement issues between the ion and water molecules occupying the space in which the ion is to be placed.

In addition to a single sodium ion, simulations using a potassium ion and magnesium ion are also performed. The environments for these simulations are created by using the same steps performed while setting up the sodium simulations.

**Results**

*Interactions between +1 ions and peptide 638 in the gas phase.*

**Sodium Ions.** Once again using data obtained from previous MD simulations, it has been seen that with gas phase MD simulations, it is possible to establish an interaction between a positively charge sodium ion and the 638 peptide fragment. With this interaction comes a stabilization effect on the alpha helical nature of the peptide. This is seen in a comparison of the two secondary structure timelines of the peptide in both systems.
Figure 44: Secondary Structure Timeline for Peptide 638 With and Without Sodium Ion

Secondary Structure timeline of the 638 Peptide without the sodium (top) and with a sodium ion nearby (bottom). Representing the stabilization effect of the ion.
**Potassium ions.** To determine that the charge specifically was the main factor resulting in the interactions between ions and peptide, another +1 charge ion was used in similar stabilization simulations. In another attempt to stabilize the alpha helical nature, as well as establish an interaction between peptide and ion, the sodium ion was replaced with a potassium ion. While the +1 charge is preserved, the size/shape of the ion differs, allowing a determination of the charge being the main culprit.

![Secondary Structure Timeline for Peptide 638 With Potassium Ion](image)

**Figure 45:** Secondary Structure Timeline for Peptide 638 With Potassium Ion MD Simulation of peptide 638 with a potassium ion at the C-terminus. Displays the same effect as a sodium ion.
Interactions between +2 ions and peptide 638 in the gas phase. Due to the nature of the interaction - positively charge ion and partial negatively charged C-terminus oxygen - the pulling force of SMD simulations allows for the extension of the peptide. The establishment of the interaction between peptide and ion becomes important. While it has been shown that positively charged ions can stabilize the alpha helical nature (sodium, potassium), if a +2 ion is used, the resulting secondary structure effects are wildly different. Instead of the stabilization of the alpha helical nature, the secondary structure is completely destroyed, not only resulting in very little alpha helical nature, but very little secondary structural features of any kind except for turns.
Figure 46: Secondary Structure Timeline for Peptide 638 With Magnesium Ion

Peptide 638 and Mg²⁺ ion (vacuum). Displays a much more destructive effect than that of the sodium ion.
It is postulated that the charge of a +2 ion in gas phase environment is much too strong to have an ordered interaction with the peptide. Instead of the ion situating into the C-terminal oxygen pocket, as seen with the +1 ions, the effect is too strong to overcome and the entire peptide "wraps around" the ion, in what seems to be a strong salting effect.
Figure 47: 3-D Representation of Peptide 638 With Potassium Ion
3-D Structure of peptide 638 interacting with the Mg$^{2+}$ ion Start of simulation (left) End of Simulation (Right)
Establishing Interactions between a sodium ion and peptide 638 in an aqueous environment. Two techniques were used to create the proper equilibrated environment. First the solvated peptide is created, then the ion is added to the aqueous environment, and then moved to the C-terminus of the peptide. The other method used the already equilibrated peptide and ion structures from gas phase environments, and then these equilibrated structures were solvated. Both techniques resulted in similar results.

Trying to establish an interaction between the ion and peptide in the aqueous phase is unclear at this point. While the secondary structure of the peptide undergoes a noticeable change in the gas phase when an interaction with an ion takes place, these changes are not visible in the peptide’s secondary structural features in the aqueous phase. As has been shown in Chapter 1, the structure of the peptide is mostly alpha helical in the gas phase, whether the ion is present or absent. However, the peptide is almost fully alpha helical in the aqueous phase. Thus, stabilization of alpha helical nature in the aqueous phase cannot be fully confirmed, since the ion stabilizes alpha helical nature, but there is no additional alpha helical nature to be gained in aqueous environments.

A further investigation is needed confirm a lack of interaction between the ion and peptide. By measuring the distance between the ion and C-terminus and tracking it, it can be determined that the ion is floating away from the peptide.
Figure 48: Distance Between the C-terminus of the Peptide and the Sodium Ion
Tracked distance between the ions and the CA of C-terminus of the 638 Peptide. Notice the Sodium ion (red line) drifts away from the peptide in the aqueous environment.
Comparing these distance results to the tracked distance of the ion and C-terminus of the peptide in the gas phase. It becomes clear that the distance should not be that great in order to allow interaction between ion and helix.

In an attempt to rectify this issue, another approach was used. First the peptide is solvated, ionized, and then the water is removed. The sodium ion is repositioned once more, but this time, instead of re-solvating and then performing an MD simulation, the peptide ion system undergoes an MD equilibration, with the resulting stable vacuum system is then solvated and used for aqueous phase MD simulations. However, the results remain the same as the first method. The solvent effect is far too strong and the ion simply floats away from the peptide.

**Using +2 ions in the aqueous phase.** Even though the gas phase simulations of the peptide and divalent ion resulted in destructive effects on the secondary structure, the solvent effect should dampen the attraction between the peptide and ion. The solvent is expected to keep the peptide from salting out the ion, instead allowing the ion to interact solely with the C-terminus, in hopes that a stabilizing interaction could be established.

The secondary structure of the peptide without the magnesium ion and with the magnesium ion are once again inconclusive, as the structure seems to go unchanged.
Figure 49: Secondary Structure Timeline of the 638 Peptide in the aqueous environments (with and without Mg ion)

Secondary Structure Timeline of the Peptide Alone in TIP3 Explicit Solvent during MD Production Simulation (Top) and Secondary Structure Timeline of the Peptide and Magnesium in TIP3 Explicit Solvent (Bottom)
However, using the tracking technique once more, it can be seen that the ion does not float away from the peptide, instead, the ion fits into the C-terminal pocket and stays put throughout the simulation.
Figure 50: Distance Between the C-terminus of the Peptide and the Mg Ion (AQ)
Magnesium ion does not drift away from the C-terminus of the peptide in the aqueous environment (blue line).
SMD Simulations

After determining that an interaction between the magnesium ion and peptide 638 was possible and once it was established, the 638 + MG system was used for SMD simulations. As previously seen in the SMD simulations of the 638 peptide and sodium ion in the gas phase, an interaction occurs between the ion and the peptide. The pull of the ion away from the origin causes the peptide to pull as well, causing an unfolding event. Eventually, the interaction between the peptide and ion will break, allowing the peptide to collapse.
Figure 51: Secondary Structure Timeline for the 638 and Mg Ion SMD Simulations (AQ)
Secondary Structure timeline resulting from SMD simulations of the 638 + MG system in a TIP3 Water box.
Conclusions

It has been shown that an interaction between +1 ions and the 638 peptide can be established in the gas phase. With this interaction comes a stabilization of the alpha-helical nature of the peptide. However, +2 ions show a destructive effect, attributed to salting out of the peptide with the ion. When trying to establish an interaction between +1 ions and the peptide in the aqueous phase, it was once again unsuccessful, as the solvation effect results in the ion floating away from the peptide before the interaction can be established. With the goldilocks effect reasoning, a +2 ion was used for aqueous environment simulations, and a successful interaction was established. With the interaction established, successful SMD simulations could be performed.

Future Simulations

The folding simulations using standard MD procedures, preserving specific hydrogen bonds, and re-folding simulations with SMD procedures yielded very promising and intriguing data. For future experimentation using these methods, experiments preserving separate individual hydrogen bonds of interest, as well as separate groups of hydrogen bonds being preserved, can be attempted in order to force the peptide to fold into a desired secondary structure. Force required to break individual hydrogen bonds can be used in an attempt to determine if the hydrogen bonds have a cumulative or singular strength attributed to each.

Finally, the continued experimentation using aqueous environments is another area of interest. While an interaction was established, and SMD simulations were
proved possible, re-folding simulations were never achieved. This is another area of interest to move toward more natural environments for peptide studies.
REFERENCES


5. Shape T. The Shape and Structure of Proteins.; 2006.


16. Pauling L. The nature of the chemical bond. {Application} of results obtained from the quantum mechanics and from a theory of paramagnetic susceptibility to the


APPENDICES

Appendices: Chapter 1

Appendix A: Ionization (Addition of a sodium ion). In order to perform the experimental simulations, a sodium ion must be added into the environment. First, the clean peptide coordinate and structure file are loaded into VMD. The VMD Solvation and Ion Modeling Extensions are utilized. The peptide system is solvated, then the addition of a single sodium ion. After this, the water is then removed from the system and the peptide and ion can be moved into the desired positions.

Appendix B: preparation and Equilibration simulations. Before any experimental simulations can be performed. A series of equilibration simulations must first be performed in order to determine energetically favorable structure of the peptide. First the structure file of the protein domain is used to extract the sequences of interest. Once only the chain of interest is obtained, it is saved as a new structure file, not containing any additional segments of the protein of no interest. The secondary structure of the static peptide can be determined using VMD's Secondary Structure Timeline feature.
Figure A 1: Secondary structure timeline for the peptide only (1 frame)

Equilibration simulations for the control simulations as well as all experimental simulations are held consistent. Once the static structure of the peptide of interest is generated, it is used as the starting structure for the control experiment. The equilibration simulation for the peptide alone last for a total of 2 ns, which consists of a 1 ns minimization period and a 1 ns equilibration simulation period. The resulting timeline is provided. This allows a "native structure" for the peptide in a gas phase environment to be obtained.
Figure A 2: Secondary Structure Timeline for the equilibration of the peptide only. (TOP) Resulting restart file (BOTTOM).

Similar to the control simulations, experimental simulations are equilibrated in a similar fashion. After the sodium ion is added to the environment, the resulting peptide and ion system is equilibrated. Again, the 2 ns total equilibration/minimization parameters are used. This allows the ion to energetically situate into a suitable position of the peptide.
Figure A 3: Secondary Structure timeline for the equilibration of the peptide and the ion.

Once stabilized, the resulting restart structure file can be used as the starting point for the experimental MD simulations. After both systems are equilibrated, the restart coordinate files can be used for all subsequent simulations.
**Figure A 4:** Use the restart for MD simulations
Failed MD Simulations:

Figure A 5: Ion Equilibration Timelines:
Equilibration of the peptide and ion system is a finicky environment. If proper equilibration is not performed, the secondary structure of the peptide can be compromised due to interactions with the ion.

It is possible for the system to become unstable and the sodium ion can pop out of place in between the carboxyl oxygen of the C-terminus.
Figure A 6: Unstable MD Simulations
Figure A 7: Timeline for the 638 Peptide when pulling the sodium ion away quickly