Abstract

NGUYEN, HUNG D. Computer Simulations of Protein Folding and Aggregation. (Under the direction of Carol K. Hall.)

Computer simulation is used to study the competition between protein folding and aggregation, especially the formation of ordered structures that are also known as amyloid fibrils. Employing simplified protein models, we simulate multi-protein systems at a greater level of detail than has previously been possible, probe the fundamental physics that govern protein folding and aggregation, and explore the energetic and structural characteristics of amorphous and fibrillar protein aggregates.

We first tackle the aggregation problem by using a low-resolution model called the lattice HP model developed by Lau and Dill. Dynamic Monte Carlo simulations are conducted on a system of simple, two-dimensional lattice protein molecules. We investigate how changing the rate of chemical or thermal renaturation affects the folding and aggregation behavior of the model protein molecule by simulating three renaturation methods: infinitely slow cooling, slow but finite cooling, and quenching. We find that the infinitely slow cooling method provides the highest refolding yields. We then study how the variation of protein concentration affects the refolding yield by simulating the pulse renaturation method, in which denatured proteins are slowly added to the refolding simulation box in a stepwise manner. We observe that the pulse renaturation method provides refolding yields that are substantially higher than those observed in the other three methods even at high packing fractions.

We then investigate the folding of a polyalanine peptide with the sequence Ac-KA$_{14}$K-NH$_2$
using a novel off-lattice, intermediate-resolution protein model originally developed by Smith and Hall. The thermodynamics of a system containing a single Ac-KA$_{14}$K-NH$_2$ molecule is explored by employing the replica exchange simulation method to map out the conformational transitions as a function of temperature. We also explore the influence of solvent type on the folding process by varying the relative strength of the sidechain’s hydrophobic interactions and backbone hydrogen bonding interactions. The peptide in our simulations tends to mimic a real polyalanine in that it can exist in three distinct structural states: $\alpha$-helix, $\beta$-structure, and random coil, depending upon the solvent conditions.

We next examine the formation of fibrillar protein aggregates, which have been implicated in the pathology of several neurodegenerative diseases including Alzheimer’s and Parkinson’s, using the Smith/Hall intermediate-resolution protein model. Simulations were conducted on systems containing 12 to 96 Ac-KA$_{14}$K-NH$_2$ peptides at a wide variety of concentrations and temperatures. We are able to observe the formation of fibrils from random coils within just a few days on a single processor of an AMD Athlon MP 2200+ workstation. To our knowledge, these are the first simulations to span the whole process of fibril formation from the random coil state to the fibril state on such a large system. We find that fibril formation strongly depends upon the peptide concentration, the temperature, and the hydrophobic interaction strength of non-polar sidechains. The fibrils observed in our simulations mimic the structural characteristics observed in experiments in that most peptides in our fibrils were arranged in an in-register parallel orientation with intra-sheet and inter-sheet distances that are similar to those observed in experiments, and are disproportionately long along the fibril axis with about six $\beta$-sheets, each of which contains many peptides.
We also investigate the kinetics of fibril formation by performing constant-temperature sim-
ulations on systems containing 48 Ac-KA_{14}K-NH_{2} peptides with the Smith/Hall intermediate-
resolution protein model. We find that fibril formation is nucleation dependent with an ordered
nucleus of two $\beta$-sheets, each with two to three peptides. The lag time before fibril forma-
tion commences decreases with increasing concentration and increases with increasing tempera-
ture. In addition, fibril formation appears to be a nucleated conformational conversion process in
which small amorphous aggregates $\rightarrow$ $\beta$-sheets $\rightarrow$ ordered nucleus $\rightarrow$ subsequent rapid growth
of a stable fibril. Fibril growth in our simulations involves both $\beta$-sheet elongation, in which the
fibril grows by adding individual peptides to the end of each $\beta$-sheet, and lateral addition, in
which the fibril grows by adding already-formed $\beta$-sheets to its side. Moreover, the rate of fibril
formation increases with increasing concentration and decreases with increasing temperature.

Finally, we examine the thermodynamics of systems containing 96 Ac-KA_{14}K-NH_{2} pep-
tides by performing replica exchange simulations over a wide range of temperatures and peptide
concentrations. We map out a phase diagram in the temperature-concentration plane delineat-
ing the regions where random coils, $\alpha$-helices, $\beta$-sheets, fibrils, and amorphous aggregates are
stable.
COMPUTER SIMULATIONS OF PROTEIN FOLDING AND AGGREGATION

by

HUNG D. NGUYEN

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

CHEMICAL ENGINEERING

Raleigh NC 27695-7905

January 9, 2004

APPROVED BY:

Carol K. Hall
Chair of Advisory Committee

Steven W. Peretti

Jason M. Haugh

Carla Mattos
This thesis is respectfully dedicated to my paternal grandmother, mè.
Biography

The author was born in Pleiku, Vietnam on April 1, 1974. He is the fifth son of Bao and Mai Nguyen and has six brothers: Dzung, Hoa, Cuong, Hiep, Thinh, and Dat Nguyen. In October 1991, he and his family migrated to the United States, relocating in Jacksonville, Florida after spending eighteen months in a refugee camp in Bataan of the Philippines. He attended Wolfson High School in Jacksonville and went to the University of Florida receiving an B.S. degree in Chemical Engineering in August, 1998. In the Fall of 1998, he was admitted to North Carolina State University to pursue graduate studies in Chemical Engineering. He received an M.S. degree in Chemical Engineering from North Carolina State University in December, 2000.
Acknowledgements

It is with pleasure that I acknowledge many of the people who have contributed their wisdom, support, and encouragement to the completion of this thesis.

I would like to thank my advisor, Professor Carol K. Hall, who has been a dedicated, enthusiastic, passionate, insightful, persistent and visionary guide in this research endeavor. Not only has she been my ever-committed technical advisor, she has also been my mentor on personal matters and career directions. I feel fortunate and honored to have worked under her guidance.

I would like to thank Dr. Anne V. Smith and Dr. Preshant Gupta for developing the models and codes that I have used in studying protein folding and aggregation. None of the work presented in this thesis would have been possible without their groundbreaking research. I also thank Drs. Hyunbum Jang, E. De Bernardez-Clark, Robert Baldwin, Sylvie Blondelle, Ken Dill, Angel García, Jeffery Kelly, Dan Kirschner, Hugh Nymeyer, and Sheena Radford for helpful discussions and advice on our studies.

I thank the National Science Foundation and the National Institutes of Health for providing generous financial support, and the High-Performance Computing center at NCSU for providing machine time, which was very much appreciated, especially during the last six months of my thesis research.

I thank the entire Hall research group, past and present, for providing a stimulating and enjoyable work environment and a warm home that is often filled with cookies and special treats. Special thanks go to Monica Hitchcock Lamm for being my confidant and role model. Particular thanks go to the group system administrators (Julie McCormick, Brian Attwood, Andrew Schultz, Aysa Akad, Arthi Jayaraman and Alex Marchut) for their incredible talent, sacrifice, and
hard work in maintaining our computer network. Extra thanks go to Andrew Schultz for helpful discussions on my entire research project and simulation codes. Also, extra thanks go to Alex Marchut for debugging parts of my code and for carrying on, along with Victoria Wagoner, our group’s protein research. I’m looking forward to learning new and exciting things about both of their projects in the near future.

I am happy to acknowledge all of the friends that I have had over the years. Special thanks go to Steven Koes, Esther Wilcox, Giles Perry, David Woodard, Doug Dankel, Randy Newswagner, Abel Bumgarner, Lee Nino, Thom Swain, and Vien Tran for their wonderful love, support and friendship. I’m lucky to have the fond memories that my friends, whom I have always considered to be my second extended family, have helped to create. Particular thanks go to J.W. Imm whose love, encouragement, support, and patience have contributed immeasurably to this work.

I wish to thank Poet John Balaban and his family for their encouragement, inspiration, generosity and particularly sacrifice in establishing the Vietnamese Nôm Preservation Foundation. I’m grateful for the books and special occasions that they have generously shared with me.

I am especially grateful for the support of my family over the duration of this thesis. I thank all of my brothers (Dzung, Hoa, Cuong, Hiep, Thinh, and Dat D. Nguyen), and their spouses or significant others for their love, encouragement and support. Special thanks go to my nephew, Anthony Khang Nguyen, who always brings smiles to everyone’s face with his infectious laughter and constant delight in discovering new things in life. I wish to thank my uncle, Quang-Anh Le, and his family for encouragement. I also wish to thank my aunt Mai, uncle John Gormican, and their extended family for their support and encouragement over the
years. Finally, I would like to close with a Vietnamese folk verse which undoubtedly does a better job than my own words ever could in expressing my gratitude to my parents, Bao and Mai Nguyen, for their sacrifice, love and support for which I will forever be thankful:

Công cha như núi Thái Sơn
Nghĩa mẹ như nước trong nguồn chảy ra
Một lòng thơ mẹ kính cha
Cho trọn chữ hiểu mối là đạo con...

– Unknown author
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiv</td>
</tr>
<tr>
<td><strong>Chapter 1  Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Overview</td>
<td>3</td>
</tr>
<tr>
<td>1.2 References</td>
<td>11</td>
</tr>
<tr>
<td><strong>Chapter 2  Effect of the Rate of Chemical or Thermal Renaturation on</strong></td>
<td>16</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Method</td>
<td>20</td>
</tr>
<tr>
<td>2.2.1 Model and sequence</td>
<td>20</td>
</tr>
<tr>
<td>2.2.2 Review of previous results by Gupta and Hall</td>
<td>21</td>
</tr>
<tr>
<td>2.2.3 Dynamic Monte Carlo simulations of four renaturation methods</td>
<td>22</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>25</td>
</tr>
<tr>
<td>2.3.1 Infinitely Slow Cooling</td>
<td>25</td>
</tr>
<tr>
<td>2.3.2 Cooling at Slow but Finite Rate</td>
<td>29</td>
</tr>
</tbody>
</table>
Chapter 3  Solvent Effects on the Conformational Transition of a Model Polyalanine Peptide

3.1  Introduction .................................................. 62
3.2  Method ...................................................... 69
   3.2.1 Model peptide and forces .................................. 69
   3.2.2 Discontinuous molecular dynamics ......................... 73
   3.2.3 Replica-exchange DMD/MC method ......................... 75
   3.2.4 Weighted histogram method ............................... 76
3.3  Results ...................................................... 79
   3.3.1 Intrinsic helical propensity of alanine ..................... 80
   3.3.2 Conformational transition as a function of temperature and hydrophobic interaction strength ..................... 83
3.4  Summary and Discussion .................................... 90
3.5  Acknowledgments ........................................... 92
3.6  References .................................................. 93
3.7  Figures ..................................................... 105
## Chapter 4  Molecular Dynamics Simulations of Spontaneous Fibril Formation from Random-Coil Peptides

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>121</td>
</tr>
<tr>
<td>4.2</td>
<td>Method</td>
<td>129</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Model peptide and forces</td>
<td>129</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Discontinuous molecular dynamics</td>
<td>131</td>
</tr>
<tr>
<td>4.3</td>
<td>Results and Discussion</td>
<td>135</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Fibril formation as a function of temperature and concentration</td>
<td>136</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Key fibril formation events</td>
<td>140</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Fibril stability</td>
<td>143</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Fibril structure</td>
<td>146</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Fibril formation as a function of the hydrophobic interaction strength between non-polar sidechains</td>
<td>149</td>
</tr>
<tr>
<td>4.4</td>
<td>Conclusions</td>
<td>150</td>
</tr>
<tr>
<td>4.5</td>
<td>Acknowledgments</td>
<td>152</td>
</tr>
<tr>
<td>4.6</td>
<td>References</td>
<td>154</td>
</tr>
<tr>
<td>4.7</td>
<td>Figures</td>
<td>170</td>
</tr>
</tbody>
</table>

## Chapter 5  Kinetics of Fibril Formation by Polyalanine Peptides

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>189</td>
</tr>
<tr>
<td>5.2</td>
<td>Method</td>
<td>195</td>
</tr>
</tbody>
</table>
5.2.1 Model peptide and forces ........................................ 195
5.2.2 Discontinuous molecular dynamics ............................... 197
5.3 Results and Discussion .................................................. 200
  5.3.1 Fibril formation is preceded by the formation of amorphous aggregates 202
  5.3.2 The rate of fibril formation is a function of peptide concentration and temperature ........................................ 205
  5.3.3 Fibril formation lag time as a function of peptide concentration and temperature ........................................ 207
  5.3.4 Fibril growth proceeds by both $\beta$-sheet elongation and lateral addition ........................................ 207
  5.3.5 Fibril formation is nucleation dependent ......................... 209
5.4 Conclusions ............................................................... 210
5.5 Acknowledgments ......................................................... 212
5.6 References ............................................................... 213
5.7 Figures ................................................................. 222

Chapter 6  Thermodynamics of Fibril Formation

by Polyalanine Peptides .................................................. 234

  6.1 Introduction ........................................................... 234
  6.2 Method ................................................................. 239
    6.2.1 Model Peptide and Forces ...................................... 239
    6.2.2 Discontinuous Molecular Dynamics ........................... 241
    6.2.3 Replica-exchange DMD/MC method ............................. 242
  6.3 Results and Discussion ............................................... 246
6.4 Conclusions .......................................................... 254
6.5 Acknowledgments ................................................... 255
6.6 References ......................................................... 256
6.7 Figures ............................................................ 264

Chapter 7 Future Work ........................................... 276

7.1 Fibril formation by alanine-based peptides ............... 276
7.2 Protein Model ...................................................... 278
7.3 Fibril formation by heteropeptides ......................... 279
7.4 References ........................................................ 283
List of Tables

<table>
<thead>
<tr>
<th>Chapter 1</th>
<th>Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2</td>
<td>Effect of the Rate of Chemical or Thermal Renaturation on the Refolding and Aggregation of a Simple Lattice Protein</td>
<td>16</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Solvent Effects on the Conformational Transition of a Model Polyalanine Peptide</td>
<td>62</td>
</tr>
<tr>
<td>3.1 Parameter $d$</td>
<td></td>
<td>72</td>
</tr>
<tr>
<td>3.2 Helix properties</td>
<td></td>
<td>82</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Molecular Dynamics Simulations of Spontaneous Fibril Formation from Random-Coil Peptides</td>
<td>121</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Kinetics of Fibril Formation by Polyalanine Peptides</td>
<td>189</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>Thermodynamics of Fibril Formation</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 7   Future Work  

276
## List of Figures

### Chapter 1 Introduction

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chapter 1 Introduction</td>
</tr>
</tbody>
</table>

### Chapter 2 Effect of the Rate of Chemical or Thermal Renaturation on the Refolding and Aggregation of a Simple Lattice Protein

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The 20-bead protein molecule in: (a) an unfolded state and (b) the native state</td>
</tr>
<tr>
<td>2.2</td>
<td>Folding transition for the model chain. The average number of HH contacts vs. $</td>
</tr>
<tr>
<td>2.3</td>
<td>Refolding yield versus $</td>
</tr>
<tr>
<td>2.4</td>
<td>Fraction proteins aggregated versus $</td>
</tr>
<tr>
<td>2.5</td>
<td>Fraction proteins unfolded versus $</td>
</tr>
<tr>
<td>2.6</td>
<td>Percentage of aggregates that are dimers and trimers versus $</td>
</tr>
</tbody>
</table>
2.7 Average times for the systems $\rho = 0.10$, 0.40 and 0.80 to equilibrate at each cooling step during the infinitely slow cooling runs (in millions of Monte Carlo timesteps). ................................................................. 54

2.8 The refolding yield at $|\epsilon^*| = 8.0$ versus the packing fraction from slow but finite cooling simulations at the cooling rates of 0.2, 1.0, 2.0, and 4.0 epsilon units per million timesteps are compared with the equilibrium refolding yield at $|\epsilon^*| = 8.0$ from the infinitely slow cooling simulations. ............................................. 55

2.9 Refolding yield versus $|\epsilon^*|$ from the quenching process at different times (in million time steps) for a system at packing fraction = 0.40; the dashed line represents data from the infinitely slow cooling case. ................................................................. 56

2.10 Percentage of chains of various types (see Figure 2.11) that participate in aggregation reactions during quenching simulations as a function of $|\epsilon^*|$. ................................................................. 57

2.11 Sample conformations illustrating the five types of HH contacts considered. ... 58

2.12 The refolding yield at $|\epsilon^*| = 6.0$ versus the packing fraction from the simulations based on suggested strategy simulations at cooling rates: 0.2, 1.0, 2.0, and 4.0 epsilon units per million time steps compared with the equilibrium refolding yield at $|\epsilon^*| = 8.0$ from the infinitely slow cooling simulations. ................................................................. 59

2.13 The refolding yield and fraction aggregated versus the packing fraction from the pulse renaturation simulations at $|\epsilon^*| = 5.0-9.0$; each addition step consisted of one chain. ................................................................. 60
2.14 The refolding yield and fraction aggregated versus the packing fraction from the pulse renaturation simulations at $|\epsilon^*| = 5.0-9.0$; each addition step consisted of five chains. ................................. 61

Chapter 3 Solvent Effects on the Conformational Transition of a Model Polyalanine Peptide 62

3.1 Geometry of the intermediate resolution protein model for polyalanine. Covalent bonds are shown with narrow solid lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_\alpha$ distances, and residue L-isomerization. Note that the united atoms are not shown full size for ease of viewing. ................................. 108

3.2 Ramachandran plots for simulations of alanine residues based on the bond length and pseudo-bond tolerance at (a) 2.0% (taken from our previous study) and (b) 2.375%. ................................. 109

3.3 Backbone hydrogen bonding. ................................. 110

3.4 Distribution of hydrogen bond (a) angles and (b) distances for $\alpha$-helices, and (c) angles and (d) distances for $\beta$-structures. ................................. 111

3.5 Snapshots of the 16-residue peptide in (a) an $\alpha$-helix and (b) a $\beta$-hairpin. Besides $\beta$-hairpins, other $\beta$-structures are $\beta$-sheet-like with (c) two $\beta$-turns, and (d) three $\beta$-turns. Hydrophobic sidechains are white, polar sidechains (at chain ends) are dark gray, and backbone atoms are light gray. Note that the united atoms are not shown full size for ease of viewing. ................................. 112
3.6 Percent $\alpha$-helical hydrogen bonds formed as a function of $T^* = k_B T/\epsilon_{HB}$ in the absence of hydrophobicity. ................................................................. 113

3.7 Percent $\alpha$-helical hydrogen bonds formed as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$. ................................................................. 114

3.8 Number of $\beta$-structure hydrogen bonds formed as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$. ................................................................. 115

3.9 Reduced potential energy of the system, $E^* = E/\epsilon_{HB}$ as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$: $R=0$ (filled circles), $R=1/12$ (open circles), $R=1/10$ (filled squares), $R=1/8$ (open squares), $R=1/6$ (filled diamonds), $R=1/4$ (open diamonds), and $R=1/2$ (filled triangles). ................................................................. 116

3.10 The conformational probability as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and $T^*$ and (b) the number of $\beta$ hydrogen bonds $n_\beta$ and $T^*$ at $R = 1/12$, $1/6$, and $1/2$. ................................................................. 117

3.11 The Helmholtz free energy of the system at $R=1/12$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures. ................................................................. 118
3.12 The Helmholtz free energy of the system at $R=1/6$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures. 119

3.13 The Helmholtz free energy of the system at $R=1/2$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures. 120

Chapter 4  Molecular Dynamics Simulations of Spontaneous Fibril Formation from Random-Coil Peptides

4.1 Covalent bonds are shown with narrow black lines connecting united atoms. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_\alpha$ distances, and residue L-isomerization. Note that the united atoms are not shown full size for ease of viewing. 173

4.2 The percentage of peptides in the $\alpha$-helical conformation obtained at the end of simulations at different temperatures and concentrations. 174

4.3 The percentage of peptides in aggregates obtained at the end of simulations at different temperatures and concentrations. 175

4.4 The percentage of peptides in $\beta$-sheet structures obtained at the end of simulations at different temperatures and concentrations. 176
4.5 The percentage of peptides in fibrillar structures obtained at the end of simulations at different temperatures and concentrations. .............................. 177

4.6 The percentage of peptides in amorphous structures obtained at the end of simulations at different temperatures and concentrations. .............................. 178

4.7 A snapshot of an aggregate obtained from a simulation at \( R=1/10, \ c=2.5\text{mM}, \) and \( T^*=0.08. \) Alpha-helices are shown in dark blue. .............................. 179

4.8 Snapshots of the 48-peptide system at various times during a simulation that results in the formation of a fibril. All hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets once the fibril is formed. .............................. 180

4.9 (a) The percentage of peptides in aggregates (of any type), \( \beta \)-sheets, and fibrillar structures and (b) the average number of peptides per \( \beta \)-sheet in a fibril and the average number of \( \beta \)-sheets per fibril versus reduced time, \( t^* \), for the 48-peptide system at \( c=20\text{mM} \) and \( T^*=0.10. \) .............................. 181

4.10 The average number of hydrogen bonds and hydrophobic interactions versus reduced time, \( t^* \), for the 48-peptide system at \( c=20\text{mM} \) and \( T^*=0.10. \) .............................. 182

4.11 Fibril stability parameters \( Q_{\text{intra-sheet}} \) and \( Q_{\text{inter-sheet}} \) versus reduced time, \( t^* \) from simulations starting with fibrils formed at \( T^*=0.11 \) performed at \( T^*=0.14 \) (filled circles), \( T^*=0.15 \) (open squares), \( T^*=0.16 \) (filled diamonds), \( T^*=0.17 \) (open triangles), and \( T^*=0.18 \) (filled triangles). .............................. 183
4.12 The average number of peptides per $\beta$-sheet in a fibril (filled circles) and the average number of $\beta$-sheets per fibril (open squares) versus reduced time, $t^*$, for the 48-peptide system at $T^*=0.15$ (top panel) and $T^*=0.16$ (bottom panel). Results are from simulations starting with fibrils formed at $T^*=0.11$ that are subsequently heated to the temperatures of interest. 

4.13 A close-up snapshot of the 48-peptide fibrillar structure viewed down the fibril axis, formed at $R=1/10$, $c=5.0\text{mM}$, and $T^*=0.13$. Note that all hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets. United atoms are not shown full size for ease of viewing.

4.14 Snapshots of fibrillar structures formed at different system sizes at $R=1/10$, $c=5.0\text{mM}$, and $T^*=0.13$.

4.15 The percentage of peptides in (a) fibrillar structures and (b) $\beta$-sheet structures at concentration $c=5\text{mM}$ versus temperature at three ratios $R$ of the strength of the hydrophobic interaction relative to that the hydrogen bond: $R=1/10$ (filled circles), $R=1/8$ (open squares), and $R=1/6$ (filled diamonds).

4.16 A snapshot of an aggregate obtained from a simulation at $R=1/6$, $c=5.0\text{mM}$, and $T^*=0.13$.

Chapter 5  Kinetics of Fibril Formation

by Polyalanine Peptides
5.1 Covalent bonds are shown with narrow black lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_\alpha$ distances, and residue L-isomerization. Note that the beads are not shown full size for ease of viewing.

5.2 Snapshots of the 48-peptide system at various times during a simulation at $c=10\text{mM}$ and $T^*=0.14$ that results in the formation of a fibril. All hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets once the fibril is formed.

5.3 (a) The percentage of peptides in aggregates (of any type), $\beta$-sheets, and fibril structures and (b) the average number peptides per $\beta$-sheet in a fibril and the average number of $\beta$-sheets per fibril versus reduced time, $t^*$, for the 48-peptide system at $c=10\text{mM}$ and $T^*=0.14$.

5.4 The percentage of peptides in aggregates (of any type), $\beta$-sheets, and fibril structures versus reduced time, $t^*$, for the 48-peptide system at a constant temperature of $T^*=0.14$ and different peptide concentrations $c=2.5\text{mM}$ (filled circles), $c=3.75\text{mM}$ (open squares), $c=5.0\text{mM}$ (filled diamonds), and $c=10.0\text{mM}$ (open triangles). 
5.5 The percentage of peptides in aggregates (of any type), β-sheets, and fibril structures versus reduced time, \( t^* \), for the 48-peptide system at a constant concentration of \( c=10.0\text{mM} \) and different temperatures \( T^*=0.13 \) (filled circles) and \( T^*=0.14 \) (open squares). ........................................ 228

5.6 The lag time for fibril formation versus the concentration for the 48-peptide system at \( T^*=0.13 \) and \( T^*=0.14 \). ................................................................. 229

5.7 The average number of peptides per β-sheet in a fibril and the average number of β-sheets per fibril versus reduced time, \( t^* \), for the 48-peptide system at a constant temperature of \( T^*=0.14 \) and different peptide concentrations \( c=2.5\text{mM} \) (filled circles), \( c=3.75\text{mM} \) (open squares), \( c=5.0\text{mM} \) (filled diamonds), and \( c=10.0\text{mM} \) (open triangles). ..................................................... 230

5.8 Schematic showing peptide locations within the fibril indicating the types of interactions (HP=hydrophobic interaction, HB=hydrogen bond) between an inner peptide, labeled “P”, and the adjacent peptides in the structure. Positions A and B represent possible extensions of the core of the fibril. ................. 231

5.9 The average number of peptides per β-sheet in a fibril and the average number of β-sheets per fibril versus reduced time, \( t^* \), from the seeded simulations of the 96-peptide system at \( c=2.0\text{mM} \) and \( T^*=0.15 \). ..................................................... 232

5.10 The percentage of peptides in fibril structures versus reduced time, \( t^* \), from the seeded and unseeded simulations of systems at \( c=2.0\text{mM} \) and \( T^*=0.14 \). ........ 233

Chapter 6  Thermodynamics of Fibril Formation

by Polyalanine Peptides
6.1 Geometry of the intermediate resolution protein model for polyalanine. Covalent bonds are shown with narrow solid lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive C\(_\alpha\) distances, and residue L-isomerization. Note that the united are not shown full size for ease of viewing.

6.2 The percentage of peptides in \(\alpha\)-helices (filled circles), fibrils (open squares), unfibrillar \(\beta\)-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=0.5\text{mM}\).

6.3 Reduced specific heat \(C_V^*\) and radius of gyration \(R_g\) (in Å) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=0.5\text{mM}\).

6.4 The percentage of peptides in \(\alpha\)-helices (filled circles), fibrils (open squares), unfibrillar \(\beta\)-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=1.0\text{mM}\).

6.5 Reduced specific heat \(C_V^*\) and radius of gyration \(R_g\) (in Å) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=1.0\text{mM}\).

6.6 The percentage of peptides in \(\alpha\)-helices (filled circles), fibrils (open squares), unfibrillar \(\beta\)-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=2.0\text{mM}\).
6.7 Reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) versus the reduced temperature $T^*$ for the 96-peptide system at $c=2.0$ mM. . . . . . . . . . . . . . . . . . 272

6.8 The percentage of peptides in α-helices (filled circles), fibrils (open squares), unfibrillar β-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature $T^*$ for the 96-peptide system at $c=5.0$ mM. . . . . . . . . . . . . . . . . . 273

6.9 Reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) versus the reduced temperature $T^*$ for the 96-peptide system at $c=5.0$ mM. . . . . . . . . . . . . . . . . . 274

6.10 Phase diagram for the 96-peptide system as a function of the reduced temperature, $T^*$, and peptide concentration, $c$. The one-structure phases are α-helices (black circles), fibrils (red circles), unfibrillar β-sheets (blue circles), and random coils (purple circles). The two-structure phases are represented by squares with the structure that has the highest percentage is listed first; they are random coils / unfibrillar β-sheets, random coils / fibrils, fibrils / unfibrillar β-sheets, α-helices / unfibrillar β-sheets, and unfibrillar β-sheets / α-helices. . . . . . . . . . . . . . . . . . 275

Chapter 7 Future Work 276
Protein molecules generally adopt a unique, three-dimensional equilibrium structure spontaneously under favorable physiological conditions $^{1,2}$. The transition of the polypeptide chain from a disordered nonnative state to the ordered, native state is called protein folding. Attempts to solve this protein folding problem - to understand how the amino acid sequence specifies the three-dimensional structure - have been underway for the past 40 years and much important progress has been made over the years.

Protein aggregation, which is often an off-pathway event associated with protein folding, is a challenging problem. It can cause economic and technical problems in the biotechnology, biomedical, and pharmaceutical industries. It can be a major nuisance in in vitro protein studies where unwanted aggregation interferes with experimental outcomes. Its effects can be lethal in patients who suffer from a variety of diseases involving protein aggregation, called amyloidoses $^{3,4}$. The most familiar amyloid diseases include Alzheimer’s, Huntington’s, Parkinson’s, and prion diseases such as the transmissible spongiform encephalopathies (best known as Mad Cow disease) and its human variant Creutzfeldt-Jacob disease. In most amyloid diseases, the aggregated protein takes the form of ordered fibrils although myositis, light-chain deposition disease and cataracts involve non-ordered protein deposits $^5$. Despite the obvious practical importance
of protein aggregation, our understanding of its physical basis is far from complete.

While experimental research is invaluable in the current effort to understand the mechanisms of protein folding and aggregation, computer simulation is a powerful complementary tool that can help to elucidate these mechanisms in molecular-level detail. Simulations can yield microscopic information about the system under investigation that cannot be obtained directly from experiments. The objective of our research is to study the competition between refolding and aggregation of proteins, especially the formation of ordered structures that are also known as amyloid fibrils. Using computer simulation, we hope to better understand how folding, aggregation kinetics, and formation of various structures are affected by the protein concentration, solvent condition, and temperature.

Given the enormous number of degrees of freedom available to a polypeptide chain, the easily-perturbed nature of the native state, and current computer limitations, one has to make a trade-off: the larger the number of degrees of freedom to be simulated, the simpler the model must be. Since our studies focus on the folding and aggregation of multi-chain systems, we employ simplified models and potentials such as the lattice HP model developed by Lau and Dill and the off-lattice, intermediate-resolution protein model originally developed by Smith and Hall. In the Lau/Dill model, each protein molecule is modeled as a sequence of hydrophobic (H) and polar (P) beads arranged in a self-avoiding random walk on a square lattice. This model incorporates the hydrophobic effect by including an attraction between any two non-bonded H beads that are situated next to each other in a chain conformation. In the Smith/Hall model, each model residue consists of a three-bead backbone and a single-bead side chain. Excluded volume, hydrogen bonds, and hydrophobic interactions are represented by discontinuous potentials. This
model is simple enough to allow the simulation of multi-protein systems over relatively long time scales yet contains enough genuine protein-like character to mimic real protein dynamics when used in conjunction with discontinuous molecular dynamics, a fast alternative to conventional molecular dynamics. Using these models allows us to sample much wider regions of conformational space, longer time scales, and larger systems than those studied by other investigators who have used low-resolution \(^{10-17}\), intermediate-resolution \(^{18-20}\), and high-resolution protein models \(^{21-35}\).

1.1 Overview

This thesis describes our research into the competition between protein folding and aggregation, especially the formation of amyloid fibrils, via molecular simulations. In this section, we summarize the remaining chapters of this thesis. All chapters contain their own literature review and bibliography. We describe a simulation study that employs the Lau/Dill lattice HP model in Chapter 2. In Chapters 3, 4, 5, and 6 of this thesis, we describe simulation studies that employ the Smith/Hall intermediate-resolution protein model. Chapter 3 describes the refinements that have been made to the Smith/Hall intermediate-resolution protein model and then discuss the folding of a single alanine-based peptide, which sets the stage for a study of polyalanine aggregation. Chapter 4 describes the formation and general features of fibrils by polyalanines. Chapter 5 examines the kinetics that are associated with fibril formation while Chapter 6 investigates the thermodynamics of fibril formation by polyalanines with the ultimate goal of mapping out a phase diagram in the temperature-concentration plane delineating the regions where different structures are stable.
In Chapter 2, we determine how variations in the rate of denaturant removal or temperature reduction affect the refolding and aggregation of the same model protein that was considered by Gupta and Hall\textsuperscript{36,37}. Three methods for the removal of denaturant are modeled: infinitely slow cooling, slow but finite cooling, and quenching. The infinitely slow cooling method, as in dialysis, requires that an infinitesimal amount of denaturant be removed at each cooling step causing the system to reach equilibrium. The slow but finite cooling study, as in multiple-step dilution or diafiltration experiments, requires that a finite fixed amount of denaturant be removed at constant time intervals. The quenching method, which is considered at much longer time scales and at a wider range of refolding conditions than Gupta and Hall, requires that a large amount of denaturant be removed instantaneously. In addition, we study the pulse renaturation method in which denatured proteins are slowly added to the refolding simulation box in a stepwise manner. In these methods, we not only monitor the average number of folded chains and aggregated chains over the course of the simulation, we also examine aggregate size. Our results show that the infinitely slow cooling method provides refolding yields that are relatively high and aggregates that are relatively small (mostly dimers or trimers). The slow but finite cooling method provides refolding yields that are almost as high as those observed in the infinitely slow cooling case but in a relatively short period of time. Quenching is extremely slow and has low refolding yields. A maximum appears in the refolding yield as a function of denaturant concentration in the simulation but disappears after a very long duration. Finally, the pulse renaturation method provides refolding yields that are substantially higher than those observed in the other three methods even at high packing fractions. As in the early stages of quenching, there is a maximum in the refolding yield as a function of denaturant concentration when a relatively high number of denatured
chains is added to the refolding solution at each step.

In Chapter 3, we describe the refinements that have been made to the Smith/Hall intermediate-resolution protein model so as to make it more suitable for the study of aggregation phenomena in very large multi-protein systems. We then perform equilibrium simulations over a very wide range of solvent conditions to determine how these affect the formation of various peptide structures. The solvent conditions considered in this study are the temperature and the hydrophobic interaction strength between non-polar sidechains. We begin by simulating the folding of an isolated model polyalanine peptide of the sequence Ac-KA_{14}K-NH_{2} over a wide range of temperatures in vacuum. The Zimm-Bragg equilibrium constants \(^ {38}\) for propagation and nucleation of the helix are calculated to quantify specific preference for the \(\alpha\)-helical state of each alanine residue on our model peptide. We then examine the effect of the hydrophobic interaction strength between non-polar sidechains on the populations of the different structural states by varying the hydrophobic interaction strength from 1/12 to 1/2 the hydrogen bonding interaction strength over a wide range of temperatures. At each value of the hydrophobic interaction strength and temperature, we calculate the conformational free energy of each structural state to determine which structure has the lowest free energy and hence locate the conformational transition between different structural states as a function of temperature. Our results show that the peptide tends to mimic a real polyalanine in that it can form different structures: \(\alpha\)-helix, \(\beta\)-structures, and random coil, depending upon the solvent conditions. The \(\beta\)-structures include \(\beta\)-hairpins and \(\beta\)-sheet-like conformations. At low values of the hydrophobic interaction strength between non-polar sidechains, the polyalanine peptide undergoes a relatively sharp transition between an \(\alpha\)-helical conformation at low temperatures and a random-coil conformation.
at high temperatures. As the hydrophobic interaction strength increases, this transition shifts to higher temperatures. Increasing the hydrophobic interaction strength even further induces a second transition to the $\beta$-hairpin state, resulting in an $\alpha$-helical conformation at low temperatures, a $\beta$-hairpin at intermediate temperatures, and a random-coil at high temperatures. At very high values of the hydrophobic interaction strength, polyalanines become $\beta$-hairpins and $\beta$-sheet-like structures at low temperatures and random-coils at high temperatures.

In Chapter 4, we investigate how peptide concentration and temperature affect the formation of various Ac-KA$_{14}$K-NH$_2$ structures including $\alpha$-helices, $\beta$-sheets and fibrils. Simulations are conducted on systems of 12, 24, 48, and 96 model 16-residue peptides at a wide variety of concentrations and temperatures. All simulations are performed in the canonical ensemble starting from a random coil configuration equilibrated at a high temperature and then slowly cooled to the temperature of interest so as to minimize kinetic trapping in local free energy minima. The percentage of peptides that form $\alpha$-helices, $\beta$-sheets or fibrils is monitored during the simulation. In addition, structural characteristics such as the peptide arrangement and packing of fibrils are examined and compared with those observed in experiments. We also study the overall stability of fibrils by conducting simulations on already-formed fibrils over a wide range of temperatures to investigate the relative importance of hydrogen bonding and hydrophobic interactions on fibril stability. The stability of our fibrillar structures is evaluated by comparing the abilities of the system to maintain the fibrillar structures at various temperatures which are higher than the fibril formation temperature. Our results show that at low concentrations, random-coil peptides assemble into $\alpha$-helices at low temperatures and random coils at high temperatures. At intermediate concentrations, random-coil peptides assemble into $\alpha$-helices at low temperatures
and large $\beta$-sheet structures at high temperatures. At high concentrations, random-coil peptides form $\beta$-sheets over a wide range of temperatures. These $\beta$-sheets assemble into fibrils above a critical temperature that decreases with peptide concentration and exceeds the isolated peptide’s folding temperature. At very high temperatures and all concentrations, the system is in a random-coil state. These results are in good qualitative agreement with the experimental results of Blondelle et al. on Ac-KA$_{14}$K-NH$_2$ peptides. The fibrils observed in our simulations mimic the structural characteristics observed in experiments in that most peptides in our fibrils are arranged in an in-register parallel orientation with intra-sheet and inter-sheet distances that are similar to those observed in experiments, and are disproportionately long along the fibril axis with about six $\beta$-sheets each of which contains many peptides. Finally, we find that when the strength of the hydrophobic interaction between non-polar sidechains relative to the strength of hydrogen bonding is high, the system forms amorphous rather than fibrillar aggregates. We also identify key fibril-forming events. Since simulations are conducted by cooling the system down to the temperature of interest, analysis of the temperature-dependence of the kinetics of fibril formation is not appropriate.

In Chapter 5, we investigate the kinetics of fibril formation of Ac-KA$_{14}$K-NH$_2$ peptides as a function of the peptide concentration and temperature. Constant-temperature simulations are conducted on systems containing 48 model 16-residue peptides at a wide variety of concentrations and temperatures using the same protein model as in our previous studies. During each simulation, the formation of different structures such as $\alpha$-helices, amorphous aggregates, $\beta$-sheets or fibrils is monitored as a function of time. Key fibril-forming events that are associated with the four proposed fibril forming mechanisms available in the literature are identified.
Two types of simulations are conducted: unseeded and seeded. In the unseeded simulations, the peptides, which are initially in a random coil configuration, are equilibrated at a high temperature and then quickly cooled to the temperature of interest. In the seeded simulations, a previously created single fibril is immersed in a sea of denatured chains. During all simulations, the lag time and rate of fibril formation are monitored to study their dependence upon the peptide concentration and temperature. Our results show that fibril formation for polyalanines incorporates features that are characteristic of three models: templated assembly, nucleated polymerization, and nucleated conformational conversion. Fibril formation is nucleation dependent with an ordered nucleus of two $\beta$-sheets, each with two to three peptides. The lag time before fibril formation commences decreases with increasing concentration and increases with increasing temperature. In addition, fibril formation appears to be a nucleated conformational conversion process in which small amorphous aggregates $\rightarrow$ $\beta$-sheets $\rightarrow$ ordered nucleus $\rightarrow$ subsequent rapid growth of a stable fibril. Fibril growth in our simulations involves both $\beta$-sheet elongation, in which the fibril grows by adding individual peptides to the end of each $\beta$-sheet and lateral addition, in which the fibril grows by adding already-formed $\beta$-sheets to its side. Moreover, the rate of fibril formation increases with increasing concentration and decreases with increasing temperature.

In Chapter 6, we perform equilibrium simulations on systems containing 96 Ac-KA$_{14}$K-NH$_2$ over a very wide range of temperatures and peptide concentrations to study the system’s thermodynamics with the ultimate goal of mapping out a phase diagram in the temperature-concentration plane delineating the regions where random coils, $\alpha$-helices, $\beta$-sheets, fibrils, and amorphous aggregates are stable. Based on the system heat capacity and peptide radius of gyra-
tion, and on the data on the percentage of the peptides that form the various structures, a phase diagram in the temperature-concentration plane is constructed. We find that there are four distinctive single-phase regions: \( \alpha \)-helices, fibrils, unfibrillar \( \beta \)-sheets and random coils and four different two-phase regions: random coils / unfibrillar \( \beta \)-sheets, random coils / fibrils, fibrils / unfibrillar \( \beta \)-sheets, and \( \alpha \)-helices / unfibrillar \( \beta \)-sheets. The \( \alpha \)-helical region is at low temperature and low concentration. The unfibrillar \( \beta \)-sheet region is at intermediate temperatures and expands to lower temperatures as concentration is increased. The fibril region is mostly at intermediate temperatures and intermediate concentrations and expands to lower and higher temperatures as the peptide concentration is increased. The random-coil region is at high temperatures and all concentrations and shifts to higher temperatures as the concentration is increased.

Chapters 2 through 6 are adapted from the following publications:


- Chapter 4: “Molecular Dynamics Simulations of Spontaneous Fibril Formation from Random-Coil Peptides”, Hung D. Nguyen and Carol K. Hall, in preparation.


- Chapter 6: “Thermodynamics of Fibril Formation by Polyalanine Peptides”, Hung D.
Nguyen and Carol K. Hall, in preparation.
1.2 References


CHAPTER 2

EFFECT OF THE RATE OF CHEMICAL OR THERMAL RENATURATION ON THE REFOLDING AND AGGREGATION OF A SIMPLE LATTICE PROTEIN

2.1 Introduction

Over-expression of recombinant proteins in prokaryotic organisms such as *Escherichia coli* often produces polypeptide chains that are accumulated in insoluble and biologically inactive aggregates called inclusion bodies. To recover active protein from inclusion bodies, the aggregates must first be isolated, then solubilized by exposure to a strong chaotropic denaturant like urea or GuHCl (guanadine hydrochloride) and finally refolded (or renatured) by removing the denaturant to the point at which the environment is conducive to folding. The rate at which the denaturant is removed has a significant effect on the refolding yield achieved. The methods currently used to remove denaturant include dilution in single or multiple steps, dialysis, and diafiltration.

Dilution is the simplest and most widely-used procedure. However, one-step dilution in which the denaturant concentration is suddenly reduced to a level that allows refolding to take
place has the disadvantage that it often leads to re-aggregation as some proteins are less soluble in the unfolded state. As a consequence, stepwise dilution to increasingly renaturing conditions is often used to improve the yield of active protein\textsuperscript{3,4}. This option works better than the one-step method because it provides enough time to destabilize the aggregates and separate chains from each other so that refolding can commence.

Dialysis and diafiltration have been shown to provide better refolding yields than dilution even at relatively high protein concentrations\textsuperscript{5–8}. Both methods employ a membrane that allows the passage of small denaturant molecules but retains the large protein molecules. As the denaturant concentration decreases and is replaced by the refolding buffer, the environment becomes more favorable for refolding. Dialysis and diafiltration have the advantage of allowing the denaturant concentration to be gradually reduced rather than to change suddenly as in one-step dilution. Dialysis has the disadvantage of being slow because it is a diffusion-limited process. Diafiltration is faster than dialysis since it is a pressure-driven process\textsuperscript{9}.

A more effective way to obtain a large amount of native protein in one operation is the method of “pulse renaturation” developed by Rudolph and Fischer\textsuperscript{10–12} or the similar “fed-batch operation” of Katoh\textit{et al.}\textsuperscript{13,14}. The renaturation process is begun at a low protein concentration. After proteins fold past a certain stage (e.g. the nucleation step) and thus become less susceptible to aggregation, another aliquot of denatured protein is added to the refolding solution. This procedure is repeated until the refolding solution becomes highly concentrated and aggregate formation threatens to dominate the refolding reaction. Pulse renaturation takes advantage of the fact that recovery of native protein from systems at low concentration of unfolded proteins is high. Adding more denatured proteins to a solution of already- or almost-folded proteins at
folding-favored conditions is in effect similar to adding denatured proteins to a solution that contains only the refolding buffer.

Although the various renaturation techniques described above often lead to improved refolding yields, our understanding of the influence of various environmental parameters such as temperature, denaturant concentration and protein concentration on aggregation is far from complete. One way to gain insight into the molecular-level mechanisms underlying protein folding and aggregation and the effect that various parameters have on the competition between protein folding and aggregation is to perform computer simulations. Interest in computer simulation of protein aggregation has grown over the past five years, due in part to ongoing improvements in computer technology and recent successes in simulating the folding of single proteins.

There have been several Monte Carlo or exact enumeration studies of the aggregation of low-resolution protein models aimed at answering some of the basic questions regarding the competition between protein folding and aggregation. Most studies were limited to only a few chains, which is not enough to fully understand the competition between folding and aggregation in a multichain system. In contrast, the systems simulated by Gupta and Hall were truly multichain systems, containing between 20 and 40 protein chains. They performed dynamic Monte Carlo simulations on a very simple two-dimensional HP lattice model of proteins to investigate the effect of solvent conditions (denaturant concentration and the protein concentration) on refolding and aggregation. They found that there were optimum levels of denaturant at which the refolding yield is highest, in agreement with some experimental observations. Although Gupta and Hall simulated their systems for long periods of time, on the order of several trillion attempted moves, their runs were not long enough, in our opinion, to reach equilibrium.
In addition, they only considered quenching processes, in which the system’s temperature or denaturant concentration is suddenly reduced to conditions that are conducive to folding. They did not explore how the rate of denaturant removal or temperature reduction affects protein refolding and aggregation.

In this paper, we determine how variations in the rate of denaturant removal or temperature reduction affect the refolding and aggregation of the same model protein that was considered by Gupta and Hall. Three methods for the removal of denaturant are modeled: infinitely slow cooling, slow but finite cooling, and quenching. The infinitely slow cooling method, as in dialysis, requires that an infinitesimal amount of denaturant be removed at each cooling step causing the system to reach equilibrium. The slow but finite cooling study, as in multiple-step dilution or diafiltration experiments, requires that a finite fixed amount of denaturant be removed at constant time intervals. The quenching method, which is considered at much longer time scales and at a wider range of refolding conditions than Gupta and Hall, requires that a large amount of denaturant be removed instantaneously. In addition, we study the pulse renaturation method in which denatured proteins are slowly added to the refolding simulation box in a stepwise manner. In these methods, we not only monitor the average number of folded chains and aggregated chains over the course of the simulation, we also examine aggregate size.

Highlights of our simulation results are the following. The infinitely slow cooling method provides refolding yields that are relatively high with only small aggregates (dimers or trimers) formed. The slow but finite cooling method provides refolding yields that are almost as high as those observed in the infinitely slow cooling case even in a relatively short time frame. Quenching, on the other hand, is relatively slow; its aggregates are comprised of more than three
molecules. A maximum in refolding yield as a function of denaturant concentration is present for the quenching case (as was observed by Gupta and Hall) but the maximum disappears after a very long simulation time. Lastly, the pulse renaturation method provides refolding yields that are substantially higher than those observed in the other methods including the infinitely slow cooling case, even at high packing fractions. When a small number of denatured chains are added at each step, the refolding yield is essentially independent of denaturant concentration. As the number of denatured chains added at each step is increased, a maximum appears in the refolding yield as a function of denaturant concentration.

2.2 Method

2.2.1 Model and sequence

The simple two-dimensional HP lattice model proposed by Lau and Dill is used to represent a protein chain. The protein molecule is a sequence of hydrophobic (H) and polar (P) beads arranged in a self-avoiding random walk on a square lattice. In order to incorporate the hydrophobic effect, two non-bonded H beads that are situated next to each other in a chain conformation attract each other with a potential of mean force of strength $\epsilon$. (Epsilon is always negative to indicate that the HH interaction is attractive.) Although the hydrophobic interaction is known to be a function of temperature, we have not taken this into account in our model for the sake of simplicity. Instead, we define the dimensionless energy parameter $\epsilon^*$ as $\epsilon^* = \epsilon/kT$ where $\epsilon$ is a constant, $T$ is the absolute temperature and $k$ is Boltzmann’s constant. In this case, the magnitude of $|\epsilon^*|$ is a measure of the denaturant concentration or temperature in the system. Low values of $|\epsilon^*|$ imply that the HH interaction is weak and hence the denaturant
concentration or temperature is high, whereas high values of $|\epsilon^*|$ imply that the HH interaction is strong and hence the denaturant concentration or temperature is low \(^{26}\). The dimensionless parameter, $|\epsilon^*|$, characterizes both inter-chain and intra-chain HH attractions. In other words, all hydrophobic contacts are treated the same regardless of whether they are inter- or intra-chain contacts \(^{26,24}\).

We focus our investigation on the same 20-bead sequence studied by Gupta and Hall: PH-PHPHPHPPHPPHPHPPHP as shown in Figure 2.1(a). This sequence was chosen for study: (1) because of its ability to fold into the native state, as displayed in Figure 2.1(b), and (2) because of its ability to form multiple, easily-recognized, partially-folded intermediates.

### 2.2.2 Review of previous results by Gupta and Hall

Since this paper builds upon the work of Gupta and Hall, it is useful to begin with a brief review of those of their results that are pertinent to the discussion here. The exhaustive conformational searches on the isolated protein chain revealed that the model molecule undergoes a relatively sharp transition as a function of $|\epsilon^*|$ from a random coil state to the native state as shown in Figure 2.2, which displays the average number of HH contacts as a function of $|\epsilon^*|$. At low values of $|\epsilon^*|$, the model chain is in an unfolded random-coil state (Figure 2.1(a)) characterized by a large radius of gyration and a small number of HH contacts. As $|\epsilon^*|$ increases, the chain folds into a compact native state (Figure 2.1(b)) with eight HH contacts arranged in a conformation with a hydrophobic core surrounded by polar beads \(^{26}\).

Dynamic Monte Carlo simulations were performed on the isolated 20-bead protein molecule in order to determine the folding pathmaps and intermediates at various levels of denat-
urant concentration. The folding process for the hypothetical protein was seen to be cooperative in that the protein chain does not start folding until the middle section of the chain is folded correctly. Once the middle section is formed, the rest of the chain zips up into the native state in a relatively short time. The model protein exhibits protein-like folding pathways and intermediates that change with the denaturant concentration.

Dynamic Monte Carlo simulations were also performed on multichain systems. These simulations were carried out in a quenching manner in that the system’s temperature or denaturant concentration was suddenly reduced to conditions conducive to folding. Aggregation was found to arise from association among partially folded intermediates, not random coil states. An optimum level of denaturant concentration where the refolding yield is highest was observed; this was attributed to the stabilization of aggregation-prone intermediates at intermediate values of the denaturant concentrations.

2.2.3 Dynamic Monte Carlo simulations of four renaturation methods

We began by examining how the variation in the rate of cooling or denaturant removal affects the refolding yield. Three renaturation processes were simulated: infinitely slow cooling, slow but finite cooling and quenching. In the infinitely slow cooling method, which mimics dialysis experiments, each simulation was started from a random initial configuration generated by growing the required number of chains in the simulation box, and then relaxing the system for a fixed amount of time at an $|e^*| = 0$ (i.e. no HH attraction) to ensure that the initial configuration was sufficiently random and hence completely denatured. At this point the HH attraction was “turned on” and the simulation started at an HH attraction strength $|e^*|=1.0$, which
is equivalent to a very high temperature or a very high denaturant concentration in experiments.

We monitored the average number of folded chains, number of aggregated chains, and total number of HH contacts in the system. A chain was defined to be folded if it formed eight intra-chain HH contacts and zero inter-chain HH contacts, aggregated if it formed at least one inter-chain HH contact, or unfolded if it formed zero inter-chain HH contacts and less than eight intra-chain HH contacts. After a long period of time (described below), the HH attraction strength $|e^*|$ was increased to 2.0. This procedure (essentially decreasing the denaturant concentration in a step-wise fashion) was repeated until $|e^*|=9.0$, which is equivalent to a very low temperature or a very low denaturant concentration in experiments. In order to determine the simulation time at each cooling step of the simulation, we divided the total simulation time at each $|e^*|$ value into four equal parts and stopped the run to switch to another condition only if the ensemble averages of the last three quarters of the run were within 2.5% of one another. Also, the ensemble averages were calculated only after each step of the simulation had completed a specified minimum number of steps. The idea here was to get the system to reach a true equilibrium state. Each system was simulated three times to obtain final values for the equilibrium refolding yield.

In the slow but finite cooling study, which mimics multiple-step dilution experiments, we simulated for a finite fixed amount of time at each $|e^*|$ before switching to the next condition using a fixed stepwise increase in $|e^*|$. (This is in contrast to the infinitely slow cooling simulations where the system equilibrates at each cooling step.) Since the cooling rate is defined as the ratio of the stepwise increase in $|e^*|$ and the (constant) simulation time at each value of $|e^*|$, we varied the cooling rate by either keeping the stepwise increase in $|e^*|$ fixed and varying the simulation time or keeping the simulation time fixed and varying the stepwise increase in $|e^*|$. At each value
of the cooling rate, each system was simulated three times to obtain the final refolding yield.

In the quenching study, which mimics one-step dilution experiments, each system was started from a random configuration equilibrated at $|\epsilon^*|=0.0$ as described above. The HH attraction term was then turned on instantaneously to the given $|\epsilon^*|$ value. The $|\epsilon^*|$ values considered here range from 1.0 to 9.0. Each system was simulated three times in order to obtain the average refolding yield.

In all of these studies, dynamic Monte Carlo simulations were performed on multichain systems containing 50 protein chains (1000 beads) at seven values of the packing fraction $\rho$ (0.10, 0.25, 0.40, 0.50, 0.59, 0.68 and 0.80) defined to be the ratio of the number of occupied sites to the total number of sites in the box. Periodic boundary conditions were used to eliminate artifacts due to simulation box wall effects. The move set employed for the simulations reported here is identical to the one used by Gupta and Hall for their aggregation studies $^{24}$; it consists of three-bond moves, four-bond moves, chain end rearrangements, and translation moves. Translational moves were rarely chosen by the simulation programs as compared to the internal moves (three bond, four bond, end rearrangements) to avoid promoting aggregation; indeed, the ratio of translational attempted moves to internal attempted moves was 0.0015. The simulation time was measured in terms of “Monte Carlo time steps”, each of which is defined as $N$ consecutive attempted moves where $N$ is the number of beads in the simulation system. Therefore, for a simulation system consisting of 50 chains with 20 beads each, one Monte Carlo step is equivalent to 1000 consecutive attempted moves.

We then studied how the variation of protein concentration affects the refolding yield by simulating the pulse renaturation method. Specifically, we studied the effect of varying the
number of new chains added at each addition step on the final refolding yield. Initially, the system was at a dilute concentration, consisting of a few chains in the denatured state ($|e^*|=0$). The HH attraction term was then turned on instantaneously to the given $|e^*|$ value. The $|e^*|$ values considered here range from 5.0 to 9.0. After a long period of time at the given $|e^*|$ value, a fresh batch of denatured chains was introduced into the simulation box. This was done by randomly picking empty lattice sites in the simulation box and growing new chains. If the growing chains got stuck, the whole system was forced to experience a Monte Carlo “move” in order to make room for the stuck chain to grow to completion. The run was stopped for another addition when the ensemble averages of the last three quarters of the run were within 2.5% of one another. This procedure of adding denatured chains to the refolding solution in steps was repeated until the final packing fraction was relatively high ($\rho=0.80$) with a total of 60 chains (1200 beads). Each system was simulated five times at a constant value of $|e^*|$ to obtain final values for the refolding yield. The addition of 1, 2, or 5 chains at each addition step was considered.

### 2.3 Results

#### 2.3.1 Infinitely Slow Cooling

The results for the infinitely slow cooling case are summarized in Figures 2.3, 2.4, and 2.5 which show the average fraction of chains that refold (refolding yield), that aggregate (fraction aggregated), and that are unfolded (fraction unfolded) as a function of denaturant concentration ($|e^*|$) for the seven above-mentioned systems. The results from three different runs at each packing fraction are within $\pm 3.10\%$ of the average value. (This is the average error over all cases considered; the largest error is $\pm 5.5\%$ and the smallest error is $\pm 2.0\%$.) For the sake of
clarity, we plot only the average values, omitting the error bars.

All the properties of interest such as the refolding yield and fractions aggregated and unfolded are highly dependent on the environment in which the system finds itself. One interesting feature that is evident in Figures 2.4 and 2.5 is that at $|\epsilon^*|=2.0$, the fraction aggregated has a maximum and the fraction unfolded has a slight minimum. As the $|\epsilon^*|$ value increases from 0.0 to 2.0, the strength of both the intra- and inter-molecular interactions increases. Intra-molecular contacts are made as some chains attempt to fold, but these chains never reach the folded state since the intra-molecular interactions are not strong enough. Inter-molecular interactions are also established; therefore, aggregation occurs. In other words, the aggregated state is preferred over the unfolded state when the native state is out of the question. When the $|\epsilon^*|$ value increases to values slightly greater than 2.0, intra-molecular interactions are strong enough to drive a small portion of the chains to the folded state at the expense of the aggregated chains; intra-molecular interactions predominate at this condition. At higher values of $|\epsilon^*|$ (6.0-9.0), folding is favored and the refolding yield reaches its maximum. Indeed, for low concentration systems, the refolding yield can reach 100%.

Once the systems attain an $|\epsilon^*|$ value of 6.0, they tend to remain stable in whatever state has been reached at this point; further increase of $|\epsilon^*|$ does not change the overall number of chains in each state. Over this range ($|\epsilon^*|=6.0-9.0$), chains are either in the aggregated or folded state, not in the unfolded state, since these two states allow all systems except those at the highest packing fractions (0.68 and 0.80) to reach their lowest energies by forming the maximum possible number of hydrophobic contacts. This does not mean that all of these systems have a final refolding yield of 100% but rather that they form some aggregates which have the maximum
possible number of hydrophobic contacts per chain. This indicates that as the conditions become more favorable for forming hydrophobic interactions, enthalpy no longer plays an important role in the competition between refolding and aggregation; these systems can reach the lowest energy regardless. Instead, the factor that decides the final number of chains that are folded or aggregated is the entropy.

The entropy is comprised of two components which are conformational and translational. The conformational component favors the formation of aggregates since the native state has only one conformation whereas aggregates are made up of chains in a variety of conformations. For example, there are at least five different conformations for a dimer that have the maximum possible number of hydrophobic contacts. The translational component of the entropy favors the formation of the native state since isolated chains are more likely to translate. At low packing fraction, the translational component dominates since most chains at low packing fractions are in the folded state. At high packing fraction, the conformational component dominates since most chains at high packing fractions are in the aggregated state. At intermediate packing fractions, the winner of the competition between the folded or aggregated states depends on the delicate interplay between these two opposing components.

The number of chains that refold or aggregate depends on the packing fraction as displayed in Figure 2.3. At low packing fractions, the refolding yield for an infinitely dilute solution ($\rho < 0.10$) is essentially the same as that of the $\rho = 0.10$ system. As the packing fraction increases beyond 0.10, the refolding yield decreases slightly at high $|\epsilon^*|$ values. As the packing fraction increases further, the refolding yield decreases more substantially at high $|\epsilon^*|$ values. This change in behavior seems to occur at a packing fraction of 0.43, a threshold region where
the model protein solution crosses over from the dilute to the semi-dilute regime as calculated by Gupta and Hall. Chains are separate in the dilute region but inter-penetrate in the semi-dilute regime. In our simulations, the protein chains are separate below a packing fraction of 0.43 and less likely to aggregate. As the packing fraction increases beyond 0.43, the protein chains begin to inter-penetrate causing a sharp increase in aggregation.

Analysis of the size of the aggregates formed during the infinitely slow cooling process indicates that aggregates are limited to either dimers or trimers as shown in Figure 2.6 which presents the average percentage of aggregates that are dimers and trimers as a function of $|\epsilon^*|$ at various values of packing fraction. At low $|\epsilon^*|$ values, this percentage has a minimum at $|\epsilon^*|=2.0$ corresponding to the condition at which the number of aggregated chains is highest for all systems as shown in Figure 2.4. At this refolding condition, most chains are in the random-coil state with many hydrophobic residues exposed, making contacts with more than three chains. As $|\epsilon^*|$ increases, the chains become more compact as most of their hydrophobic residues become buried in the interior. At high $|\epsilon^*|$ values (6.0-9.0) all systems except those at $\rho=0.68$ and 0.80 form aggregates that are comprised of dimers or trimers with the maximum number of hydrophobic contacts per chain.

The time that each system takes to equilibrate at each cooling step depends on the value of $|\epsilon^*|$ and $\rho$. Figure 2.7 presents the average time to equilibrate as a function of the denaturant concentration for three systems ($\rho=0.10$, 0.40 and 0.80) during the infinitely slow cooling simulations. Here time is measured in terms of “Monte Carlo time steps” which consists of 20 consecutive attempted moves made by each chain. The two ends of the $|\epsilon^*|$ value range are very stable in that each system takes a relatively short time to equilibrate; the middle range is less
stable in that each system takes much longer time to equilibrate and the systems at high packing fractions require especially long times. The middle region corresponds to the folding transition region $3.5 \leq |\epsilon^*| \leq 5.5$. Within this region, the ensemble averages of the folded and aggregated chains fluctuate wildly before they settle down to their equilibrium value.

We have also simulated infinitely slow cooling with a stepwise increase in the $|\epsilon^*|$ value of 0.5 (data not shown) instead of 1.0 as reported in this section. The final refolding yield, fraction aggregated, and simulation times at each cooling step are the same (within experimental error) as those obtained from the infinitely slow cooling simulations with stepwise increases of $\epsilon$ equal to 1.0. This suggests that for infinitely slow cooling, smaller stepwise increases of the $|\epsilon^*|$ value are unnecessary; bigger stepwise increases in $|\epsilon^*|$ would require a much longer time to reach equilibrium as occurs in the quenching case, this will be discussed in a later section.

The data that are reported here are taken to be equilibrium data because all systems were simulated for a very long period of time and their results adequately satisfied the equilibrium criteria that the ensemble averages vary no more than 2.5% during the last three quarters of the run. In fact, at low temperatures ($|\epsilon^*| > 6.0$), the ensemble averages varied by much less than $\pm 2.5\%$. Moreover, the ensemble averages did not change at all once the system reached $|\epsilon^*|=6.0$. Thus we call these simulations infinitely slow cooling runs.

### 2.3.2 Cooling at Slow but Finite Rate

Experiments on protein refolding do not usually involve infinitely slow cooling or infinitely slow denaturant removal. Instead, experimental methods (such as multiple-step dilution, dialysis, and diafiltration) require that the denaturant be removed from the refolding solution either in
discrete steps or continuously. This means that the time interval for each cooling step is fixed. Here we study the effects of varying the rate of removing denaturant (or reducing temperature) on the final refolding yield.

Slow but finite cooling runs were conducted in which the $|\epsilon^*|$ value was increased by 1.0 in a stepwise fashion for simulation times of 0.25, 0.50, 1.00 and 5.00 million time steps. This is equivalent to cooling rates of 4.0, 2.0, 1.0, and 0.2 epsilon units per million time steps. These simulation times are relatively short compared to the infinitely slow cooling case whose shortest cooling interval lasted 10 million steps. Figure 2.8 reports the refolding yield versus the packing fraction at these different cooling rates. The top line represents equilibrium data taken at $|\epsilon^*|=8.0$; each subsequent line represents data obtained at the end, i.e. at $|\epsilon^*|=8.0$, of a slow but finite cooling run at the specified rate. As evident in this figure, the cooling rate is important; increasing the cooling rate decreases the refolding yields. Low cooling rate provides refolding yields that are relatively close to the equilibrium yield but are 10 times quicker than that of the infinitely slow cooling case; higher cooling rate provides refolding yields that deviate more strongly from the equilibrium data. We also studied the effect of cooling rate by fixing the simulation time at one million time steps for each cooling step and considering four different stepwise increases in the $|\epsilon^*|$ value: 0.2, 1.0, 2.0, and 4.0 (Data not shown). Since the cooling rates were the same as in the first case (0.2, 1.0, 2.0, and 4.0 epsilon units per million time steps), the two studies gave identical results. It is apparent that renaturation can be accomplished through either long time intervals between steps or removal of little denaturant at each step.
2.3.3 Quenching

Here, we report our studies of the effects of quenching on the refolding yield over a much longer time scale and wider range of refolding conditions than Gupta and Hall. Recall that they found that there were optimum levels of denaturant at which the refolding yield was highest, in agreement with some experimental observations\(^5,32^-{35}\). Gupta and Hall stopped their simulations when the ensemble averages of folded chains, aggregated chains, and HH contacts in the system during the last third of each simulation were within 2.5\% of the ensemble averages for the middle third of the simulation. We wondered if this stopping criterion was sufficient to ensure equilibrium because the ensemble averages could change after a long period of time. In the current investigation, we tested whether or not further simulation could erase the refolding maximum as has been reported experimentally by De Bernardez-Clark \(^{36}\), who found that the maximum in the refolding of hen egg white lysozyme as a function of guanidinium chloride concentration vanishes after a long period of time.

A maximum in the refolding yield versus denaturant concentration, similar to that reported by Gupta and Hall, does appear in our studies of quenching. This is displayed in Figure 2.9, which shows how the refolding yield obtained from the quenching method varies with \(|\epsilon^*|\) at various times during the simulation (0.1, 10, 80, 160, and 320 million time steps) for a system at \(\rho=0.40\). Early in the renaturation process, systems at \(|\epsilon^*|\) values \(\leq 5.0\) reach equilibrium quickly and then fluctuate about equilibrium whereas systems at higher \(|\epsilon^*|\) values reach equilibrium more slowly. A sharp maximum in the refolding yield appears at \(|\epsilon^*|=5.0\) after 0.1 million time steps and remains there after 10 million timesteps. The maximum seems to broaden and shift slightly to the right as the length of the run increases; for instance, after 80 million time
steps, the maximum widens as the system at $|e^*|=6.0$ reaches the same refolding yield as the one at $|e^*|=5.0$. Similarly, after 160 million, the system at $|e^*|=7.0$ reaches the same refolding yield as that at $|e^*|=5.0-6.0$. Indeed, even after 320 million time steps, which is relatively long compared to the 45 million time steps in the slow but finite cooling case, the systems at $|e^*|=8.0$ and 9.0 have not achieved the same refolding yield as those at $|e^*|=5.0-7.0$. As indicated in this figure, the systems at $|e^*|=6.0-9.0$ did not reach equilibrium. The refolding yield at $|e^*|=6.0-9.0$ would reach equilibrium after billions of timesteps, but this would take months of CPU time on our 500-MHz Alpha workstations. Essentially, as time increases the maximum broadens as its peak shift toward higher $|e^*|$ values. After a long period of time, the maximum should flatten out over the higher end of $|e^*|$ values and cease to exist. We believe this to be the case because the equilibrium line, which is equivalent to an infinitely slow cooling run, displays no maximum.

Analysis of the aggregates during the quenching process indicates that they are relatively stable at high $|e^*|$ values. Indeed, each chain in an aggregate forms more than seven hydrophobic contacts on average after 320 million time steps (data not shown). However, these aggregates do not form the highest number of interactions possible as occurred in the infinitely slow cooling case. In addition, the aggregates are bigger than those observed in the infinitely slow cooling case. As time increases, each chain in an aggregate experiences an increased number of interactions, making the aggregates even more stable. Releasing a chain from an aggregate so that the refolding reaction can take place increases the system’s energy, which is unfavorable. In fact, the chance that such an event takes place is very slim; therefore, the rate of de-aggregation and thus refolding is extremely small.

Most of the chains that become aggregated contain a relatively small number of non-native
hydrophobic contacts. This is shown in Figure 2.10, which represents the probability that a loose chain prior to aggregation possesses a certain type of HH contact as a function of $|\epsilon^*|$. Each point represents the average value of data over ten simulation runs on the seven systems ($\rho = 0.10, 0.25, 0.40, 0.50, 0.59, 0.68$ and $0.80$). The five types of HH interactions referred to in Figure 2.10 are illustrated in Figure 2.11. Type I represents local native HH contacts between either of the last two hydrophobic residues at the end of the chain arms (beads 2 and 5 or beads 16 and 19). Type II represents local native HH contacts in the middle of the chain arms (beads 5 and 8 or beads 13 and 16). Type III represents local native HH contacts at the center of the chain (beads 8 and 13 or beads 9 and 12). Type IV represents non-local native HH contacts between the end of the chain arms and the chain middle (beads 2 and 9 or beads 12 and 19). Type V represents non-local, non-native HH contacts between the chain ends (beads 2 and 13, 2 and 19, 5 and 9, 5 and 12, 5 and 16, 8 and 19). Thus, all types except Type V represent HH contacts that are found in the native state. As indicated in Figure 2.10, aggregation results from association between almost-folded intermediates rather than between random coil proteins. Chains containing local HH contacts (represented by Types I, II, and III) have a high probability of forming aggregates at high $|\epsilon^*|$ values. Chains containing non-local interactions (represented by Types IV and V) have a lower probability of forming aggregates. Chains containing non-native interactions, which are represented by Type V, have the lowest probability of forming aggregates. This indicates that aggregation reactions involve intermediates with substantial native-like content.

The critical role of the rate of denaturant removal in determining the refolding yield for systems at high protein concentration has been observed experimentally for the different renaturation methods. Yoshii et al. studied the refolding of 5 mg/ml lysozyme using the diafiltration
method (equivalent to the slow but finite cooling case of our simulations). They found that when the urea concentration was decreased from an initial value of 8M at a rate of 2.5 M/hr, the refolding yield was 40%. When the denaturant-removal rate was 1.26 M/hr, the refolding yield was 60% and when the rate was 0.48 M/hr, the refolding yield reached a maximum at 75%. They also found that when the urea concentration was decreased from an initial value of 4M (which is still a high denaturant concentration) at a very slow rate of 0.167 M/hr, the refolding yield was 85% after 11 hours. These findings seem to agree qualitatively with the results from our slow but finite cooling case as displayed in Figure 2.8, which shows that the refolding yield for a system at $\rho=0.40$ consistently improves (by as much as 12%) as the cooling rate is reduced from 4.0 to 0.2 epsilon units per million time steps. At a higher lysozyme concentration of 10 mg/ml, Yoshii et al. found that the refolding yield was reduced to 68%, which agrees with the trend in our studies (lower refolding yields at high protein concentrations). Maeda et al. observed a similarly high refolding yield (82%) at a lysozyme concentration of 5 mg/ml by using the dialysis method (equivalent to the infinitely slow cooling case of our simulations). This method took a longer period of time (66 hours) than Yoshii’s method (11 hours), which agrees with our observation that the slow but finite cooling case can provide refolding yields that are almost as high as those in the infinitely slow cooling case but in a shorter period of time. Using a conventional rapid dilution method (equivalent to the quenching case of our simulations), De Bernardez-Clark et al. observed an 82% refolding yield in 24 hours at a lysozyme concentration of 5 mg/ml when the denaturant guanidinium chloride was withdrawn. This required a longer time than Yoshii’s diafiltration experiments to get to a similar refolding yield, which agrees with our finding that a longer time is needed to obtain high refolding yields by the quenching method than by the
slow but finite cooling method. De Bernardz-Clark’s dilution method required a shorter time than Maeda’s dialysis method because De Bernardz-Clark et al. searched for and found an optimum set of conditions at which high refolding yield could be obtained over a relatively short period of time. This agrees with our finding that there exists a range of time-dependent optimum conditions (at the middle $|\epsilon^*|$ values) in the quenching case at which the refolding yield is highest.

The dependence of thermal renaturation on cooling rate is similar to that of chemical renaturation on denaturant-removal rate. The effects of both slow and rapid cooling on aggregation were examined by Tani et al.\textsuperscript{38,39}. They investigated the folding of ovalbumin that was heat-denatured at 80 degrees C and then cooled either slowly or rapidly. When cooled slowly, the heat-denatured ovalbumin refolded to its native structure. But when cooled rapidly the heat-denatured molecules were forced into metastable non-native conformations which tended to aggregate for several days before transforming slowly to the native conformation. This finding seems to agree with our studies in the sense that slow cooling destabilizes aggregates by forming a few small aggregates whereas quenching promotes the formation of numerous large aggregates early in the renaturation process.

Based on what we have learned from the simulations of the infinitely slow cooling, slow but finite cooling and quenching cases, we suggest a strategy for obtaining high refolding yields in a relatively short time. From the quenching simulations, we learned that all our systems can reach equilibrium very quickly at low $|\epsilon^*|$ values, say up to 4.0. From the infinitely slow or slow but finite cooling studies, we learned that cooling is unnecessary at $|\epsilon^*|$ values above 6.0. This suggests that high refolding yields could be obtained for our systems in a short time by
quenching at $|\epsilon^*|=4.0$, and then slow cooling to $|\epsilon^*|=6.0$. In order to test this idea, the system was quenched from $|\epsilon^*|=0.0$ to 4.0 and then cooled slowly as in the slow but finite cooling method to $|\epsilon^*|=6.0$. The stepwise increase in the $|\epsilon^*|$ value was fixed at 1.0 and the simulation time for each run was 0.25, 0.50, 1.00 and 5.00 million time steps, which is equivalent to cooling rates of 4.0, 2.0, 1.0, and 0.2 epsilon units per million time steps. Figure 2.12 reports the refolding yield observed at $|\epsilon^*|=6.0$ versus the packing fraction at these different cooling rates. The top line represents equilibrium data taken at $|\epsilon^*|=8.0$ from the infinitely slow cooling case; each subsequent line represents data obtained at $|\epsilon^*|=6.0$ at the specified rate of slow but finite cooling from the quenched configuration at $|\epsilon^*|=4.0$. As is evident in this figure, low cooling rate also provides refolding yields that are relatively close to the equilibrium yield as in the slow finite cooling case. Thus, quenching from $|\epsilon^*|=0.0-4.0$ and then cooling slowly to $|\epsilon^*|=6.0$ provides the same refolding yields as if the system was cooled in a slow but finite fashion from $|\epsilon^*|=0.0-8.0$ but is three times quicker.

2.3.4 Pulse Renaturation

Here we describe our attempts to simulate pulse renaturation. Initially, the system contained a few chains in the denatured state. After the refolding yield and fraction aggregated remained constant, a fresh batch of denatured chains was introduced into the simulation box. This procedure of adding denatured chains to the refolding solution in steps was repeated until the final protein concentration reached $\rho=0.80$. Three cases were considered: addition of 1, 2, or 5 chains at each step. These systems were simulated at constant values of $|\epsilon^*|$ equal to 5.0, 6.0, 7.0, 8.0, and 9.0. Each step lasted 32 million timesteps for the one-chain addition runs, 90M for the
two-chain addition runs, and 120M for the five-chain addition runs.

The refolding yields are very high, especially when the rate of addition is one chain per addition step. This is shown in figures 2.13 and 2.14, which represent the refolding yield and fraction aggregated versus the instantaneous value of the packing fraction after adding one chain per step (Fig. 2.13) and five chains per step (Fig. 2.14) at five values of $|\epsilon^*|$. The refolding yields are even higher than those of the infinitely slow cooling case. For example, at packing fractions greater than 0.50, the refolding yield obtained from this method is twice as high as that observed in the infinitely slow cooling simulations shown in Figure 2.3. The packing fraction does not have as much of an impact on the refolding yield in this case as it does in the infinitely slow cooling case; when the packing fraction increases from 0.40 to 0.80, the refolding yield in pulse renaturation decreases at most by 30% (for $|\epsilon^*|=5.0$) as compared to 65% in the infinitely slow cooling case.

As is evident in Figure 2.13, refolding yields at the highest values of $|\epsilon^*|$ are insensitive to $|\epsilon^*|$. In other words, folding yields at $|\epsilon^*|=6.0-9.0$ are essentially identical when each addition step consists of one chain. The same is true for runs in which each addition step consists of two chains (data not shown), but the refolding yields when each addition step consists of five chains are an average of 10% lower than the refolding yields when each addition step consists of one chain.

As the number of chains added per addition step increases, the refolding yield depends more sensitively on the value of $|\epsilon^*|$ as indicated by Figure 2.14. In addition, there is an intermediate range of $|\epsilon^*|$ at which the refolding yield is highest. The refolding yields at $|\epsilon^*|=6.0-7.0$ are higher than those at $|\epsilon^*|=5.0$, 8.0 and 9.0 over the whole range of packing fractions. This
maximum is similar to that observed during the early period of the quenching simulations.

Even though the pulse renaturation systems were simulated five times at each $|e^*|$ value and the same stopping criteria as in the infinitely slow cooling case was used, there is still considerable scatter in the resulting averages of the refolding yield and the fraction aggregated as seen in the two figures. During some addition steps, the refolding yield actually improved as more denatured chains were added. This could happen when the system has not had a chance to equilibrate yet; addition of new chains could further perturb the system causing some chains to go on to the native state. The reason could be that the simulations were done at high $|e^*|$ values (low temperatures) at which the systems were so frozen that they would require much longer simulation times to register a change and that 2.5% might be too high as an equilibrium criteria. Thus, the pulse renaturation results do not appear to be at equilibrium.

2.4 Discussion

We examined the refolding and aggregation kinetics of the multiprotein lattice model introduced by Gupta and Hall. We looked at how the rate of changing solvent concentration and temperature affect folding and aggregation by modeling four renaturation methods - infinitely slow cooling, slow but finite cooling, quenching, and pulse renaturation. Our studies seem to be able to predict the trends observed experimentally. Quenching is detrimental to high refolding yields; aggregates are formed early in the renaturation process and become more or less frozen, minimizing the number of available isolated chains for refolding. Both the infinitely slow and the slow but finite cooling methods, however, promote high refolding yields; aggregates are destabilized at an early stage of the renaturation process, maximizing the number of available unfolded
and isolated chains for refolding. Thus, renaturation via either infinitely slow cooling or slow but finite cooling method is more effective than quenching; i.e., it can provide higher refolding yields in a shorter period of time at any packing fraction. However, the effectiveness of all of the methods considered depends strongly on the protein concentration. In the infinitely slow and slow but finite cooling cases, there is a range of low protein concentrations at which the refolding yield can approach 100%; the refolding yield decreases rapidly as the protein concentration increases beyond this range. At high protein concentration, the pulse renaturation method is superior to the three other methods considered provided that the rate of adding denatured chains is low and the denaturant concentration is neither too high or too low.

It is important to point out that our model and analysis are subject to a number of limitations. The model is extremely simple in that it is two-dimensional and does not account for the presence of solvent molecules explicitly. Like all two-dimensional HP lattice models, the model does not satisfy the experimental criterion for calorimetric two-state cooperativity \(^{40}\). Applying the two-state calorimetric criteria to the collapse transition of our HP model protein using previously determined specific heat and energy data \(^{26}\), the enthalpy ratio was calculated to be 0.56 by Chan’s method \(^{40}\) and 0.32 by Zhou’s method \(^{41}\). These values are significantly smaller than unity and thus do not satisfy the two-state calorimetric criteria, indicating that our HP model does not represent the thermodynamics of real proteins in detail. Our choice of move set is not as realistic as it could be. It could lead to nonergodicity problems or possible bias toward the native state because it does not include nonlocal moves such as rigid rotations \(^{42-44}\). These potential problems are not as likely to occur in our multichain simulations as they would be for isolated systems for the following two reasons. First, rigid rotation moves are not likely to succeed at
high packing fractions. Secondly, our end rearrangement moves in which the two end beads on the chain were randomly reoriented are particularly effective at unfolding the chains at high packing fractions even under strongly native conditions. In fact, folded chains in our simulations had a high probability (0.10) of breaking two out of eight native hydrophobic contacts and in the process exposing three hydrophobic residues to the surroundings. Since end rearrangements involve only two beads at a time, they have a higher probability of being accepted in tightly packed systems than rigid rotation moves which often involve more than two beads at a time. Thus, possible biases toward the native state, as opposed to the aggregated or unfolded state, were minimized. Nevertheless, we should point out that despite the model’s limitations, it has allowed us to delineate the general trends that are observed in experiment.

2.5 Acknowledgments

This work was supported by the National Institutes of Health (Grant No GM-56766) and the National Science Foundation (Grant No CTS-9704044). The authors wish to acknowledge useful discussions with E. De Bernardez-Clark.
2.6 References


[34] M. Goldberg, *In Dynamic aspects of conformation changes in biological macromolecules* (Reidl. publ. dordrecht, Netherlands, 1982).


2.7 Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The 20-bead protein molecule in: (a) an unfolded state and (b) the native state.</td>
<td>48</td>
</tr>
<tr>
<td>2.2</td>
<td>Folding transition for the model chain. The average number of HH contacts vs. $</td>
<td>\epsilon^*</td>
</tr>
<tr>
<td>2.3</td>
<td>Refolding yield versus $</td>
<td>\epsilon^*</td>
</tr>
<tr>
<td>2.4</td>
<td>Fraction proteins aggregated versus $</td>
<td>\epsilon^*</td>
</tr>
<tr>
<td>2.5</td>
<td>Fraction proteins unfolded versus $</td>
<td>\epsilon^*</td>
</tr>
<tr>
<td>2.6</td>
<td>Percentage of aggregates that are dimers and trimers versus $</td>
<td>\epsilon^*</td>
</tr>
<tr>
<td>2.7</td>
<td>Average times for the systems $\rho = 0.10, 0.40$ and 0.80 to equilibrate at each cooling step during the infinitely slow cooling runs (in millions of Monte Carlo timesteps).</td>
<td>54</td>
</tr>
<tr>
<td>2.8</td>
<td>The refolding yield at $</td>
<td>\epsilon^*</td>
</tr>
</tbody>
</table>
2.9 Refolding yield versus $|e^*|$ from the quenching process at different times (in million time steps) for a system at packing fraction = 0.40; the dashed line represents data from the infinitely slow cooling case. .......................... 56

2.10 Percentage of chains of various types (see Figure 2.11) that participate in aggregation reactions during quenching simulations as a function of $|e^*|$. .............. 57

2.11 Sample conformations illustrating the five types of HH contacts considered. . . . . 58

2.12 The refolding yield at $|e^*| = 6.0$ versus the packing fraction from the simulations based on suggested strategy simulations at cooling rates: 0.2, 1.0, 2.0, and 4.0 epsilon units per million time steps compared with the equilibrium refolding yield at $|e^*| = 8.0$ from the infinitely slow cooling simulations. .............. 59

2.13 The refolding yield and fraction aggregated versus the packing fraction from the pulse renaturation simulations at $|e^*| = 5.0-9.0$; each addition step consisted of one chain. ................................................................. 60

2.14 The refolding yield and fraction aggregated versus the packing fraction from the pulse renaturation simulations at $|e^*| = 5.0-9.0$; each addition step consisted of five chains. ................................................................. 61
Figure 2.1: The 20-bead protein molecule in: (a) an unfolded state and (b) the native state, taken from 26.
Figure 2.2: Folding transition for the model chain. The average number of HH contacts vs. $|\epsilon^*|$, taken from $^{26}$. 
Figure 2.3: Refolding yield versus $|\epsilon^*|$ from the infinitely slow cooling simulations for systems at $\rho = 0.10, 0.25, 0.40, 0.50, 0.59, 0.68$ and $0.80$. 
Figure 2.4: Fraction proteins aggregated versus $|\varepsilon^*|$ from the infinitely slow cooling simulations for systems at $\rho = 0.10, 0.25, 0.40, 0.50, 0.59, 0.68$ and $0.80$. 
Figure 2.5: Fraction proteins unfolded versus $|\epsilon^*|$ from the infinitely slow cooling simulations for systems at $\rho = 0.10, 0.25, 0.40, 0.50, 0.59, 0.68$ and $0.80$. 
Figure 2.6: Percentage of aggregates that are dimers and trimers versus $|\epsilon^*|$ from the infinitely slow cooling simulations for systems at packing fractions $\rho = 0.10, 0.25, 0.40, 0.50, 0.59, 0.68$ and 0.80.
Figure 2.7: Average times for the systems $\rho = 0.10$, 0.40 and 0.80 to equilibrate at each cooling step during the infinitely slow cooling runs (in millions of Monte Carlo timesteps).
Figure 2.8: The refolding yield at $|\epsilon^*| = 8.0$ versus the packing fraction from slow but finite cooling simulations at the cooling rates of 0.2, 1.0, 2.0, and 4.0 epsilon units per million timesteps are compared with the equilibrium refolding yield at $|\epsilon^*| = 8.0$ from the infinitely slow cooling simulations.
Figure 2.9: Refolding yield versus $|\epsilon^*|$ from the quenching process at different times (in million time steps) for a system at packing fraction = 0.40; the dashed line represents data from the infinitely slow cooling case.
Figure 2.10: Percentage of chains of various types (see Figure 2.11) that participate in aggregation reactions during quenching simulations as a function of $|\epsilon^*|$. 
Figure 2.11: Sample conformations illustrating the five types of HH contacts considered.
Figure 2.12: The refolding yield at $|\epsilon^*| = 6.0$ versus the packing fraction from the simulations based on suggested strategy simulations at cooling rates: 0.2, 1.0, 2.0, and 4.0 epsilon units per million time steps compared with the equilibrium refolding yield at $|\epsilon^*| = 8.0$ from the infinitely slow cooling simulations.
Figure 2.13: The refolding yield and fraction aggregated versus the packing fraction from the pulse renaturation simulations at $|\epsilon^*| = 5.0-9.0$; each addition step consisted of one chain.
Figure 2.14: The refolding yield and fraction aggregated versus the packing fraction from the pulse renaturation simulations at $|\epsilon^*| = 5.0-9.0$; each addition step consisted of five chains.
Chapter 3

Solvent Effects on the Conformational Transition of a Model Polyalanine Peptide

3.1 Introduction

Small peptides undergo a spontaneous reversible disorder-to-order transition to a unique, three-dimensional equilibrium structure such as an $\alpha$-helix or a $\beta$-structure when exposed to favorable physiological conditions. The information necessary to drive this folding transition is universally accepted to be encrypted solely within the linear amino acid sequence. However, experiments indicate that many peptides can be folded into alternative stable structures by changing the solution conditions. For example, the conformational transition between the $\alpha$-helix and $\beta$-structure is greatly influenced by solvent conditions including the pH, the temperature, the salt or organic cosolvent concentration, the peptide concentration, and the redox state. Moreover, the $\alpha$-$\beta$ conformational transition is believed to play an important role in the assembly of normally soluble proteins into ordered aggregates, known as amyloid fibrils, which are a cause or associated symptom of many human disorders, including Alzheimer’s and the prion diseases. In some of these disorders, an $\alpha$-helix-rich protein
slowly accumulates into long, straight and unbranched ribbons of layered crossed $\beta$-sheets\textsuperscript{25–30}, destroying the architecture and function of the surrounding tissue, usually with degenerative and ultimately-fatal consequences. In addition, recent evidence shows that proteins other than those associated with amyloid diseases form fibrils \textit{in vitro} under mildly denaturing conditions\textsuperscript{31–33}. Despite the important role played by the solvent conditions in the $\alpha$-$\beta$ transition, our understanding of these effects is far from complete. Here computer simulations are performed in an attempt to examine the effect of solvent conditions on the $\alpha$-$\beta$ transition of model polyalanine peptides.

The effect of solvent conditions on the thermodynamics and kinetics of polyalanine and alanine-based peptides has been examined via computer simulations with atomistic models by a number of workers. One of the earliest simulations of the complete helix-coil transition was a study by Daggett \textit{et al.}\textsuperscript{34} who performed molecular dynamics simulations of the denaturation of a 13-residue polyalanine $\alpha$-helix using the forcefield by Levitt \textit{et al.}\textsuperscript{35} and the program ENCAD\textsuperscript{36}. They found that the helix is stable at all temperatures ranging from 5 to 200 degree C in vacuum but only at low temperatures in solution. Vila \textit{et al.}\textsuperscript{37} performed conformational Monte Carlo searches on ALA10 and ALA16. They observed that the helicity for both peptides was close to zero when the model dielectric constant was 80, which corresponds to the value for pure water, and was over 90% when the model dielectric constant was 2, which corresponds to the value within a protein environment without solvent polarization effects. Hansmann, Okamoto, and coworkers conducted multicanonical simulations using the forcefield ECEPP/2\textsuperscript{38} over a wide range of temperatures, and found that the helix-coil transition of polyalanines of lengths 10 to 30 residues is a true thermodynamic phase transition in vacuum\textsuperscript{39–42}. The transition temperature was significantly lower in water, implying that presence of solvent tends to weaken helix
formation \cite{43-45}. García et al.\cite{46,47} investigated the thermodynamics of the helix-coil transition of a 21-residue polyalanine peptide in explicit water by conducting replica-exchange molecular dynamics (REMD) simulations with a modified AMBER force field. They found that the helix-coil transition is moderately cooperative over a broad temperature range. Ohkubo et al.\cite{48} examined equilibrium helix-coil transitions of Ac-Ala$_n$-NMe ($n = 3$-20) as a function of temperature using REMD with the CHARMm/GB implicit solvent force field. They found that the intrinsic helical propensity of alanine measured by the Zimm-Bragg parameters is independent of the chain length. Olivella et al.\cite{49} performed nanosecond molecular dynamics simulations using the AMBER95 forcefield of Cornell et al.\cite{50} on polyalanine $\alpha$-helices in polar and non-polar solvents such as water and methane, respectively. They found that the intra-helical hydrogen bond is weaker in polar solvent than in non-polar solvent. In contrast, a study by Mortenson et al.\cite{51}, who performed all-atom simulations on ALA12 and ALA16 using the same forcefield as Olivella et al.\cite{49}, showed that changing the environment from in vacuo to aqueous conditions increases helicity.

All of the simulation studies mentioned above predict a transition between the $\alpha$-helix and random coil but not the existence of a $\beta$-structure. Two recent studies, however, do predict the formation of a $\beta$-structure by polyalanines \cite{52,53}. Levy et al.\cite{52} performed molecular dynamics simulations using the CHARMm program on polyalanine ALA12 at several temperatures, and found that the native state is an $\alpha$-helix in vacuum and a $\beta$-structure in aqueous solution. Ding et al.\cite{53} performed molecular dynamics simulations by applying a modified version of our implicit-solvent protein model \cite{54,55} to DeGrado and coworkers’ $\alpha_{1B}$ amphipathic $\alpha$-helix peptide \cite{56} and observed the formation of a $\beta$-hairpin. Based on their calculation of the free energy of the
\(\beta\)-hairpin, they concluded that the \(\beta\)-hairpin is a metastable intermediate. Since Ding et al.\textsuperscript{53} conducted constant-temperature simulations at low temperatures, their system might have been trapped in local minima, which would mean there was insufficient sampling of those parts of configurational space that are important in thermodynamic calculations.

In this paper, we perform equilibrium simulations over a very wide range of solvent conditions to determine how these affect the formation of various peptide structures. The solvent conditions considered in this study are the temperature and the hydrophobic interaction strength between non-polar sidechains. We begin by simulating the folding of an isolated model polyalanine peptide over a wide range of temperatures in vacuum; in this case, the hydrophobic interaction is turned off. We examine the effect of temperature on the populations of the peptide’s various structural states and locate the conformational transitions between different structural states. The Zimm-Bragg equilibrium constants\textsuperscript{57} for propagation and nucleation of the helix are calculated to quantify specific preference for the \(\alpha\)-helical state of each alanine residue on our model peptide. We then examine the effect of the hydrophobic interaction strength between non-polar sidechains on the populations of the different structural states by varying the hydrophobic interaction strength from \(1/12\) to \(1/2\) the hydrogen bonding interaction strength over a wide range of temperatures. At each value of the hydrophobic interaction strength and temperature, we calculate the conformational free energy of each structural state to determine which structure has the lowest free energy and hence locate the conformational transition between different structural states as a function of temperature.

The model peptide chosen for study is the polyalanine-based peptide \textit{Ac-KA}_{14}\textit{K-NH}_{2}. We focus on polyalanine-based peptides for three reasons. First, the small, uncharged, unbranched
nature of alanine residues is amenable to simulation with the intermediate-resolution protein model that we developed previously\textsuperscript{54,55}. Second, polyalanine repeats have been implicated in human pathologies; in particular, they are responsible for the formation of anomalous filamentous intranuclear inclusions in oculopharyngeal muscular dystrophy patients\textsuperscript{58}. Third, synthetic polyalanine-based peptides have been shown by Blondelle \textit{et al.} to undergo a transition from $\alpha$-helical structures to $\beta$-sheet complexes at concentrated conditions \textit{in vitro}\textsuperscript{59,60}, mimicking the structural transition believed to be a prerequisite for fibril nucleation and growth\textsuperscript{61–67}. Blondelle \textit{et al.} observed that the $\alpha$-helical structures were stabilized in part by intramolecular $\alpha$-helical bonds and that the macromolecular $\beta$-sheet complex was stabilized by hydrophobic inter-sheet interactions. Using circular dichroism, fourier-transform infrared spectroscopy, and reversed phase high performance liquid chromatography, they found that: (1) $\beta$-sheet complex formation increased with increasing temperature, exhibiting an S-shaped dependence on temperature with a critical temperature of 45°C at a peptide concentration of 1.8mM and an incubation time of 3h, and (2) $\beta$-sheet complex formation increased with increasing peptide concentration above a critical concentration of 1mM at 65°C.

The polyalanine-based peptide Ac-KA\textsubscript{14}K-NH\textsubscript{2} is represented using an off-lattice, implicit-solvent, intermediate-resolution protein model originally introduced by Smith and Hall\textsuperscript{54,55,68}. In this paper, we describe refinements to this model to make it more suitable for the study of aggregation phenomena in very large multi-protein systems. Three modifications have been made to the Smith/Hall model. Firstly, the constraint on the fluctuation of the bond length between neighboring united atoms is relaxed by increasing its tolerance from ±\textsuperscript{2.0%} about the ideal value to ±\textsuperscript{2.375%} about the ideal value; this produces a more realistic Ramachandran plot than
was previously obtained by extending the $\Phi$-$\Psi$ space explored to include $\Phi$-$\Psi$ angles that correspond to the left-hand $\alpha$-helix and polyproline II $^{69}$. Increasing the bond length fluctuation tolerance also increases the code’s speed since the united atoms move a greater distance between collisions. Secondly, the hydrogen bond formation algorithm was modified to include a penalty for breaking the bond when the hydrogen bond angle is too extreme. Thirdly, the efficiency techniques that have been developed in our group for modeling polymer dynamics are implemented, which increases the code’s speed by a factor of 50. These modifications are discussed in greater detail in the Methods section.

Simulations are performed using the discontinuous molecular dynamics (DMD) simulation algorithm $^{70–73}$, which is an extremely fast alternative to traditional molecular dynamics and is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials. Unlike soft potentials such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide, enabling the exact (as opposed to numerical) solution of the collision dynamics. DMD simulations proceed by locating the next collision, advancing the system to that collision, and then calculating the collision dynamics. In order to ensure that our simulations have the capacity to effectively sample configurational space without getting trapped in local minima, we modify the replica-exchange method $^{74}$ for use with our DMD algorithm simulation. Here we follow the approach of Sugita and Okamoto $^{75}$, who originally formulated replica exchange as a combined molecular dynamics and Monte Carlo (MD/MC) method. In replica-exchange, multiple replicas of the same system are simulated over a wide range of temperatures. At set time intervals, replicas whose temperatures are nearest neighbors along the temperature spectrum are exchanged provided that a Metropolis cri-
tion is satisfied. This procedure is repeated until all of the systems at different temperatures reach equilibrium. Once at equilibrium, the data collection phase begins in which the probability of various energy levels and states are stored for use in a weighted histogram calculation \(^{76,77}\) of the conformational free energy of the various peptide structures. This simulation procedure is discussed in greater detail in the Methods section.

Highlights of our simulation results are the following. We find that peptides in our simulations tend to mimic real polyalanines in that they can exist in three distinct structural states: \(\alpha\)-helices, \(\beta\)-structure ensemble (including \(\beta\)-hairpin and \(\beta\)-sheet-like structures), and random coils, depending upon the solvent conditions. At low values of the hydrophobic interaction strength between non-polar sidechains, the polyalanines undergo a relatively sharp transition between an \(\alpha\)-helical conformation at low temperatures and a random-coil conformation at high temperatures. As the hydrophobic interaction strength increases, this transition shifts to higher temperatures. Increasing the hydrophobic interaction strength even further induces a second transition to a \(\beta\)-hairpin, resulting in an \(\alpha\)-helical conformation at low temperatures, a \(\beta\)-hairpin at intermediate temperatures, and a random-coil at high temperatures. At very high values of the hydrophobic interaction strength, polyalanines become \(\beta\)-hairpin and \(\beta\)-sheet-like structures at low temperatures and random-coils at high temperatures. This study of the folding of a single polyalanine-based peptide sets the stage for a study of polyalanine aggregation in a forthcoming paper.

This paper is organized as follows. Section 2 describes the methods used in this work, including the protein’s physical representation, its potential energy function, the DMD simulation technique, the replica-exchange method, and the weighted histogram method. Section 3 presents
the results obtained from simulations at various conditions. Section 4 contains a summary and
discussion of our findings and a brief description of our future work on fibril formation in multi-
peptide systems based on the sequence studied in this paper.

3.2 Method

3.2.1 Model peptide and forces

The model peptide is 16 residues long with the sequence PH\textsubscript{14}P, where H stands for a
hydrophobic amino acid residue and P stands for a polar amino acid residue. This sequence was
chosen to approximate Ac-KA\textsubscript{14}K-NH\textsubscript{2} peptides which have been shown by Blondelle et al.\textsuperscript{59,60}
to form stable, soluble $\beta$-sheet complexes. The peptide is represented at an intermediate level of
resolution using a model introduced by Smith and Hall\textsuperscript{54,55,68}. Details of the model including
values for all parameters are given in earlier papers. The model is based on a four-bead amino
acid representation with realistic bond lengths and bond-angle constraints and has the ability
to interact both intra- and inter-molecularly via hydrogen bonding and hydrophobic interaction
potentials. The geometry of the protein model is illustrated in Figure 3.1. Each amino acid
residue is composed of four spheres, a three-sphere backbone comprised of united atom NH,
C\textsubscript{\alpha}H, and C=O, and a single bead sidechain R (these are labeled N, C\textsubscript{\alpha}, C, and R, respectively
in the figure). All backbone bond lengths and bond angles are fixed at their ideal values; the
distance between consecutive C\textsubscript{\alpha} atoms is fixed so as to maintain the interpeptide bond in the
trans configuration. The sidechains are held in positions relative to the backbone such that all
residues are L-isomers.

The solvent is modeled implicitly in the sense that its effect is factored into the energy
function as a potential of mean force. All forces are modeled by either hard-sphere or square-well potentials. The excluded volumes of the four beads in each residue are modeled using hard-sphere potentials with realistic diameters. Covalent bonds are maintained between adjacent spheres along the backbone by imposing hard sphere repulsions whenever the bond lengths attempt to move outside of the range between \((l-\delta)\) and \((l+\delta)\) where \(l\) is the bond length and \(\delta\) is a tolerance. Ideal backbone bond angles, \(C_{\alpha}-C_{\alpha}\) distances and residue L-isomerization are achieved by imposing pseudobonds as shown in Figure 3.1, which also fluctuate within a tolerance \(\delta\). In previous studies \(^{54,55,68}\), the bond length and pseudobond tolerance \(\delta\) were fixed at 2.0\%, producing a relatively restricted region of sterically allowed \(\Phi-\Psi\) space confined to the left side of the Ramachandran plot as shown in Figure 3.2(a) \(^{54}\). In the present study, the bond length and pseudobond tolerance \(\delta\) are relaxed to a value of 2.375\%, producing a more realistic Ramachandran plot. This is seen in Figure 3.2(b), which shows a plot of \(\Phi\) versus \(\Psi\) from our simulation on a peptide with small, alanine-sized side chain residues. The sterically allowed region of \(\Phi-\Psi\) space on the left side of the Ramachandran plot is larger than in our previous paper. The shape of this region agrees better with standard, published data that describe the regions of sterically allowed \(\Phi-\Psi\) space in real proteins \(^{69}\). All secondary structures including polyproline II and the left-hand \(\alpha\)-helix (right side of plot) \(^{69}\) are observed in these simulations.

Hydrogen bonding between amide hydrogen atoms and carbonyl oxygen atoms on the same or neighboring chains is represented by a square-well attraction, whose depth is \(\epsilon_{HB}\) and range is \(1.5\sigma_R\) where \(\sigma_R\) is the sidechain bead diameter between NH and C=O united atoms whenever: (1) the virtual hydrogen and oxygen atoms (whose location can be calculated at any time) are separated by 4.2Å (the sum of the NH and C=O well widths), (2) the nitrogen-hydrogen and
carbon-oxygen vectors point toward each other within a fairly generous tolerance, (3) neither the
NH nor the C=O are already involved in a hydrogen bond with a different partner, and (4) the
NH and C=O are separated by at least three intervening residues along the chain. Smith and Hall
satisfied the second requirement above by determining the locations of the virtual oxygen
and hydrogen atoms based on the vector projections as shown by the arrow tips in Figure 3.3
and by stipulating that both the NHO and the COH angles be between 120° and 180°. In
this paper, we adopt an approach similar to that of Ding et al. and require that the four atom
pairs $N_i-C_{α,j}$, $N_i-N_{j+1}$, $C_j-C_{α,i}$, $C_j-C_{i-1}$ shown connected by thick dashed lines in Figure 3.3
(hereafter referred to as auxiliary pairs), be separated by distances $d$ that are chosen to maintain
the hydrogen bond angle constraints; their values are given in Table 3.1. Upon the formation of
a bond between $N_i$ and $C_j$, these auxiliary pairs temporarily interact via a single-step square-
shoulder potential:

$$u_{ij}(r) = \begin{cases} ∞, & r ≤ σ \\ ε_{HB}, & σ < r ≤ d \\ 0, & r > d \end{cases} \quad (3.1)$$

where $r$ is the distance between beads $i$ and $j$, $σ$ is the bead diameter, $ε_{HB}$ is the shoulder height
which is equal to the well depth of the hydrogen bond between $N_i$ and $C_j$, and $d$ is the square-
shoulder width. This interaction is more efficient computationally than that of Ding et al. who
used a two-step square-shoulder potential to maintain hydrogen bond angles. Besides preventing
the hydrogen bond angles from straying outside of the desired values, the square-shoulder
interaction within each auxiliary pair adds stability to the hydrogen bond by increasing the range
of the repulsive forces between the beads that neighbor $N_i$ and $C_j$ as was done previously by
Table 3.1: Parameter $d$

<table>
<thead>
<tr>
<th>Pairs</th>
<th>$d$ (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_i-C_{\alpha,j}$</td>
<td>5.00</td>
</tr>
<tr>
<td>$N_i-N_{j+1}$</td>
<td>4.74</td>
</tr>
<tr>
<td>$C_{j-C_{\alpha,i}}$</td>
<td>4.86</td>
</tr>
<tr>
<td>$C_{j-C_{i-1}}$</td>
<td>4.83</td>
</tr>
</tbody>
</table>

Takada, Luthey-Schulten, and Wolynes \cite{78} and by Smith and Hall \cite{54,55,68}. These auxiliary pairs return to their original interactions when the hydrogen bond is broken.

Interactions between hydrophobic sidechains are represented by a square-well potential of depth $\epsilon_{HP}$ and range $1.5\sigma_R$ where $\sigma_R$ is the sidechain bead diameter. Hydrophobic sidechains must be separated by at least three intervening residues in order to interact. The strengths of the hydrophobic contact, $\epsilon_{HP}$, considered in this paper are 0, 1/12, 1/10, 1/8, 1/6, 1/4, and 1/2 the strength of a hydrogen bond, $\epsilon_{HB}$. Hydrogen bond strength and hydrophobic contact strength are independent of temperature, as has been assumed in previous simulation studies \cite{79,55,68,80}.

In our simulations, the criteria above lead to hydrogen bonds with realistic $OH$ distances and $NHO$ angles as illustrated in Figure 3.4, which shows the distributions of the $COH$ and $NHO$ angles and the $OH$ distances for both $\alpha$-helices and $\beta$-structures. Snapshots of an $\alpha$-helix and different $\beta$-structures are projected as Rasmol \cite{81} renderings in Figure 3.5. The results in Figure 3.4 agree better with the hydrogen bond properties reported by Baker and Hubbard based on an extensive analysis of $\alpha$-helices and $\beta$ structures in a high-resolution protein database \cite{82} than those reported in our previous study \cite{54}. For our $\alpha$-helices, the $NHO$ angle varies between 115 and 180 degrees with a mean of 148 degrees and the $COH$ angle varies between 110 and 170 degrees with a mean of 131 degrees. In comparison, Baker et al. report a $NHO$ angle ranging from 110 to 180 degrees with a mean of approximately 157 degrees and a $COH$ angle ranging
from 110 to 180 degrees with a mean of approximately 147 degrees. For our $\beta$-structures, the $NHO$ angle varies between 120 and 180 degrees with a mean of 160 degrees and the $COH$ angle varies between 110 and 180 degrees with a mean of 154 degrees. In comparison, Baker et al. report a $NHO$ angle ranging from 120 to 180 degrees with a mean of approximately 162 degrees and a $COH$ angle ranging from 120 to 180 degrees with a mean of approximately 151 degrees. For our $\alpha$-helices, the hydrogen bond length (virtual hydrogen - to - virtual oxygen distance) varies between 1.8 and 2.8Å with a mean of 2.23Å. This distribution is somewhat different from the range (1.6-2.5 Å) and mean (2.04Å) reported by Baker and Hubbard. For our $\beta$-structures, the hydrogen bond length varies between 1.8 and 2.9Å with a mean of 2.06Å. This distribution is also somewhat different from the range (1.6-2.5 Å) reported by Baker and Hubbard but in better agreement with their mean (1.94Å).

### 3.2.2 Discontinuous molecular dynamics

Simulations are performed using the discontinuous molecular dynamics (DMD) simulation algorithm $^{70-73}$, which is an extremely fast alternative to traditional molecular dynamics and is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials. DMD simulations are conducted as follows. Each bead of the model protein chain is assigned a random initial position and a random initial velocity that do not violate any of the size constraints or assigned bond lengths and angles. The initial velocities are chosen at random from a Maxwell-Boltzmann distribution at a specified reduced temperature $T^* = k_B T / \epsilon_{HB}$, where $k_B$ is Boltzmann’s constant, $T$ is temperature, and $\epsilon_{HB}$ is the strength of the hydrogen bond in the model as explained earlier. When a DMD simulation begins, each
bead moves with its individual velocity. The simulation proceeds according to the following schedule: identify the first event, move forward in time until that event occurs, calculate new velocities for the pair of beads involved in the event and calculate any changes in system energy resulting from hydrogen bond events or hydrophobic interactions, find the second event, and so on. Types of events include excluded volume events, bond events, and square-well hydrogen bond and hydrophobic interaction events. An excluded volume event occurs when the surfaces of two hard-sphere beads collide and repel each other. A bond (or pseudobond) event occurs via a hard-sphere repulsion when two adjacent spheres attempt to move outside of their assigned bond length. Square-well events include well-capture, well-bounce, and well-dissociation “collisions” when a sphere enters, attempts to leave, or leaves the square well of another sphere. For more details on DMD simulations with square-well potentials, see papers by Alder and Wainwright \(^{70}\) and Smith, Hall, and Freeman \(^{83}\).

Simulations are performed in the canonical ensemble which means that the number of particles, volume, and temperature are held constant. Periodic boundary conditions are used to eliminate artifacts due to simulation box wall effects. The dimensions of the box are chosen to ensure that a chain cannot interact with more than one image of any other chain. For this study, we use a cubic box with sides 100Å in length. Constant temperature is achieved by implementing the Andersen thermostat method \(^{84}\) as was used previously \(^{54,77}\). With this procedure, all beads in the simulation are subject to random collisions with ghost particles. The post-event velocity of a bead colliding with a ghost particle is chosen randomly from a Maxwell-Boltzmann distribution at the simulation temperature.
3.2.3 Replica-exchange DMD/MC method

The replica-exchange method is implemented with 26 replica systems distributed over a broad interval of temperature ranging from $T_{*}=0.08$ to a high temperature at which the peptide is a total random coil. Each replica system is simulated at a different temperature $T$ in the canonical ensemble using the DMD method. The number of replicas and the distribution of temperatures are chosen to ensure that: (a) there is a free random walk in temperature space, which means that every replica has the same probability of being switched to a neighboring temperature; (b) the number of replicas and hence temperatures sampled must be high enough to ensure that the probability of each replica being switched to a neighboring temperature is greater than 10%; and (c) the highest temperature sampled must be high enough to prevent the system from becoming trapped in a local energy minimum. These requirements are the same as those stated by Sugita and Okamoto. At fixed time intervals, replicas are sorted from lowest to highest temperature and subjected to the following temperature exchange MC procedure. Systems $i$ and $j$, with neighboring temperatures $T_i$ and $T_j$, respectively, can exchange configurations (system $i$ changes to temperature $T_i$ and system $j$ to temperature $T_j$) with probability:

$$
Probability = \begin{cases} 
1 & \text{if } \Delta \leq 0 \\
\exp(-\Delta) & \text{if } \Delta > 0
\end{cases}
$$

(3.2)

where $\Delta = [\beta_j - \beta_i](U_i - U_j)$ with $\beta_i = 1/(k_BT_i)$ and $U_i$ the potential energy of the system in state $i$. Initially each system is in a random configuration obtained from an NVT simulation at high temperature. Exchange attempts occur every 20.0 reduced time units where the reduced time is defined as $t^* \equiv t/\sigma\sqrt{k_BT/m}$ where $t$ is the simulation time, and $\sigma$ and $m$ are the average
bead diameter and mass; this corresponds to approximately five million collisions at each low
temperature and six million collisions at each high temperature. Approximately one thousand
replica-exchange attempts are made during our simulations before equilibrium is reached. Our
equilibrium criteria requires that the ensemble averages of the system’s total potential energy,
which is collected at the end of each DMD run, vary by no more than 2.5% during the second
half of all DMD runs at each temperature. Once equilibrium is reached, the data collection
phase begins in which three hundred extra replica-exchange attempts are made. During the data
collection phase, the properties of interest at each temperature are calculated throughout each
DMD run. In addition, the probability of being in different energy levels and states is stored
for conformational free energy calculations of the different thermodynamic properties of interest
using the weighted histogram method. At the end of the replica-exchange DMD/MC simulation,
our data contain a large ensemble of peptide configurations at each temperature. Our simulations
last for a long time taking more than six billion collisions at each low temperature and eight
billion collisions at each high temperature. A replica-exchange DMD/MC simulation at a single
hydrophobic interaction strength requires 25 days on a single processor of an AMD Athlon MP
2200+ workstation.

3.2.4 Weighted histogram method

The conformational free energy of the various structures of the peptide is calculated using
the weighted histogram method\textsuperscript{76,77} in order to determine the thermodynamic stability of each
structure at different solvent conditions. The weighted histogram method is a least-squares op-
timization method that allows the calculation of degeneracy factors from simulations at several
temperatures which can then be used to calculate the thermodynamic properties of interest. In this method, each peptide conformation (state) is expressed in terms of the number of $\alpha$-helical hydrogen bonds $i$, the number of $\beta$-structure hydrogen bonds $j$, and the number of hydrophobic interactions $k$. The partition function $Z$ can then be expressed as a sum over the energy levels of the various states $(i,j,k)$:

$$Z(T_l) = \sum_{i} \sum_{j} \sum_{k} g_{i,j,k} e^{(-\beta_l E_{i,j,k})}$$

(3.3)

where $\beta_l=1/k_B T_l$ and $E_{i,j,k}$ and $g_{i,j,k}$ are the energy and degeneracy factor for the energy levels in energy state $(i,j,k)$. Details of our modified weighted histogram method are as follows. During the replica-exchange simulation, the probability $P_{i,j,k}(T_l)$ of being in energy state $E_{i,j,k}$ at temperature $(T_l)$ is monitored at each temperature $T_l$. At the end of the simulation, the probabilities $P_{i,j,k}(T_l)$ can be normalized as:

$$\sum_{i} \sum_{j} \sum_{k} P_{i,j,k}(T_l) = 1$$

(3.4)

where $I=12, J=7$, and $K=18$ are the maximum possible number of $\alpha$-helical hydrogen bonds, $\beta$-structure hydrogen bonds and hydrophobic interactions, respectively. The degeneracy factor, $g_{i,j,k}$, from simulations at all of the temperatures considered $(T_l, l=1,2,...,L)$ can be calculated as:

$$g_{i,j,k} = \frac{\sum_{l=1}^{L} P_{i,j,k}(T_l)}{\sum_{l=1}^{L} e^{-\beta_l E_{i,j,k} - J_l}}$$

(3.5)
where \( f_l \) is an optimization parameter which is related to the Helmholtz free energy and partition function at temperature \( T_l \) via \( f_l = -\beta_l A(T_l) = \ln Z(T_l) \). It can be expressed as:

\[
f_l = \ln(\sum_i \sum_j \sum_k g_{i,j,k}e^{-\beta_l E_{i,j,k}})
\] (3.6)

The degeneracy factor \( g_{i,j,k} \) and the \( f_l \) for \( l=1,2,\ldots,L \) can be calculated by solving the last two equations iteratively with simulation data \( P_{i,j,k}(T_l) \) as input. Once initial guesses for the \( f_l \) at each \( T_l \) are made (\( f_l=0 \) is used), \( g_{i,j,k} \) can be calculated from equation 3.5, which is in turn used to calculate \( f_l \) from equation 3.6. If the difference between the new and old values of all of the \( f_l \) are small, then the iteration has converged and the newly calculated \( g_{i,j,k} \) and \( f_l \) are accepted. Otherwise, the iteration continues as above using the new values of \( f_l \) as the initial guesses until the iteration converges. The conformational free energy in each energy state at temperature \( T_l \) can be calculated from the resulting \( g_{i,j,k} \) as

\[
A_{i,j,k}(T_l) = E_{i,j,k} - \frac{\ln(g_{i,j,k})}{\beta_l}
\] (3.7)

One can use the conformational free energy \( A_{i,j,k}(T_l) \) to compare the free energy of the various energy states. To facilitate the comparison, the conformational free energy \( A_{i,j,k}(T_l) \) in each energy state \((i,j,k)\) is broken down into two terms, \( A_{i,j}(T_l) \) and \( A_{m,k}(T_l) \). \( A_{i,j}(T_l) \) is the conformational free energy associated with the numbers of \( \alpha \)-helical, \( i \), and \( \beta \)-structure hydrogen bonds, \( j \), regardless of the number of hydrophobic interactions, \( k \). \( A_{m,k}(T_l) \) is the conformational free energy associated with the total number of hydrogen bonds, \( m \), and the number of hydrophobic interactions, \( k \). Here \( m \) is the sum of the number of \( \alpha \)-helical and \( \beta \) hydrogen
bonds \((m = i + j)\) with the possible maximum number of hydrogen bonds being \(M=12\). The quantities \(A_{i,j}(T_i)\) and \(A_{m,k}(T_i)\) can be calculated once the numerical procedure described in the previous paragraph is complete. First, the probability \(P_{i,j,k}^*(T_i)\) of being in energy state \(E_{i,j,k}\) is reconstructed (hence referred to as the reconstructed probability, \(P_{i,j,k}^*(T_i)\)) from the calculated \(A_{i,j,k}(T_i)\) since \(P_{i,j,k}^*(T_i) \propto e^{\beta_i A_{i,j,k}(T_i)}\) and then normalized at each temperature. The final value for the conformational free energy \(A_{i,j}(T_i)\) is then:

\[
A_{i,j}(T_i) = -\frac{1}{\beta_i} \ln(\sum_{k=0}^{K} P_{i,j,k}^*(T_i)) 
\]  

(3.8)

Similarly, the conformational free energy \(A_{m,k}(T_i)\) can be expressed as:

\[
A_{m,k}(T_i) = -\frac{1}{\beta_i} \ln(\sum_{i=0}^{I} \sum_{j=0}^{J} P_{i,j,k}^*(T_i)) 
\]  

(3.9)

In the Results section, \(A_{i,j}(T_i)\) and \(A_{m,k}(T_i)\) are presented in isosurface plots to compare the conformational free energy of the various structures at different temperatures.

### 3.3 Results

In this section, we present results from replica-exchange DMD/MC simulations on the structures formed by the isolated, 16-residue model KA14K peptide. One structure of interest is a full-length \(\alpha\)-helix with four \(\alpha\)-helical turns, twelve \(\alpha\)-helical hydrogen bonds (defined as bonds between \(N_{i+4}\) and \(C_i\)), and ten hydrophobic contacts. Other structures of interest are those that have a significant number of \(\beta\) hydrogen bonds. A \(\beta\) hydrogen bond is defined to be a hydrogen bond between two residues whose backbone angles are in the \(\beta\)-region of the Ramachandran plot.
(0° ≤ ψ ≤ 180° and -180° ≤ φ ≤ -30°). In this paper, β-structures are defined to be structures having three or more β hydrogen bonds. Examples include β-hairpins and β-sheet-like structures. We focus particularly on the β-hairpin, which is defined to a structure having one β-turn and six or more consecutive β hydrogen bonds which link one half of the peptide to the other half. We also refer to some structures as being β-sheet-like by which we mean a structure that has two or more β-turns, three to five β hydrogen bonds. Snapshots of an α-helical structure, β-hairpin, and two β-sheet-like structures for the model peptide are shown in Figure 3.5 with hydrophobic residues colored white, polar residues colored dark gray, and backbone atoms colored light gray. Note that the united atoms are not shown full size for ease of viewing.

3.3.1 Intrinsic helical propensity of alanine

In the absence of hydrophobic interactions, model alanine residues on the KA_{14}K peptide have a strong tendency to form α-helical hydrogen bonds. This is apparent from the high number of α-helical hydrogen bonds formed at low temperatures as shown in Figure 3.6, which displays the normalized number of α-helical hydrogen bonds (relative to the perfect α-helix value of 12) at various reduced temperatures (T* = k_B T/ε_{HB}) averaged over a large ensemble of the peptide configurations at each temperature collected during the replica-exchange simulations when ε_{HP} = 0. Indeed, the alanine residue’s specific preference for the α-helical state is confirmed by calculating the parameters s (equilibrium constant for propagation of the helix) and σ_{ZB} (nucleation of the helix) in the Zimm-Bragg model. According to this model, the propagation parameter, s, which is a measure of helix propensity, can have a value of one (the chain is 50% helical and the helix and coil are equally favored), greater than one (the helix is favored relative to a random
coil), and less than one (the helix is disfavored relative to a random coil). An increase in $s$ corresponds to increases in both the average number of helical residues ($< n >$) and the average length of a helical segment ($< l_{ZB} >$). The nucleation parameter, $\sigma_{ZB}$, which is a measure of the propensity for nucleating a helical segment, can have values between zero and one, with a value of zero corresponding to a large barrier to nucleation and a value of one corresponding to no barrier to nucleation. As $\sigma_{ZB}$ increases towards a value of one, the increased number of nucleation events corresponds to the creation of more helical segments, $< n >$, which has the effect of lowering the average segment length, $< l_{ZB} >$. For more details on the calculation of the Zimm-Bragg parameters, see our previous simulation study. Table 3.2 shows $s$ and $\sigma_{ZB}$ along with the average number of helical residues, $< n >$, and average length of a helical segment, $< l_{ZB} >$, observed in our simulations at different reduced temperatures, $T^*$. At low temperatures, high values of $s$ (1.80) indicate that the alanine residues on the model peptide have a strong preference to form helix over coil, which is in good agreement with preferences observed in experiment and in the simulations of Okamoto and Hansmann, Garciá and Sanbonmatsu, and Ohkuho and Brooks.

The specific preference for the $\alpha$-helical state of alanine residues by the model peptide depends on the temperature as indicated by Table 3.2. The propagation parameter, $s$, decreases with increasing temperature, ranging from its highest value of $s=1.80$ at $T^*=0.08$, meaning that the helix is favored relative to a random coil, to its lowest value of $s=0.24$ at $T^*=0.1042$, meaning that the helix is disfavored relative to a random coil. The transition temperature at which $s=1$, where the chain is 50% helical, is between $T^*=0.0986$ and $T^*=0.0994$. The $s$ values in our study are in good agreement with experimental results which yield estimates of $s$ for alanine between
Table 3.2: Helix properties measured during simulations (average number of α-helical residues, \( < n > \) and average length of an α-helical segment, \( < l_{ZB} > \)) and calculated Zimm-Bragg parameters (propagation parameter, \( s \), and nucleation parameter, \( \sigma_{ZB} \)) for different values of reduced temperature.

<table>
<thead>
<tr>
<th>( T^* )</th>
<th>( &lt; n &gt; )</th>
<th>( &lt; l_{ZB} &gt; )</th>
<th>( s )</th>
<th>( \sigma_{ZB} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0800</td>
<td>11.96 ± 0.01</td>
<td>11.95 ± 0.02</td>
<td>1.80</td>
<td>0.09</td>
</tr>
<tr>
<td>0.0940</td>
<td>11.73 ± 0.03</td>
<td>11.57 ± 0.06</td>
<td>1.65</td>
<td>0.08</td>
</tr>
<tr>
<td>0.0954</td>
<td>10.53 ± 0.32</td>
<td>10.31 ± 0.34</td>
<td>1.28</td>
<td>0.04</td>
</tr>
<tr>
<td>0.0970</td>
<td>10.35 ± 0.35</td>
<td>9.82 ± 0.40</td>
<td>1.26</td>
<td>0.05</td>
</tr>
<tr>
<td>0.0986</td>
<td>8.08 ± 0.54</td>
<td>7.85 ± 0.55</td>
<td>1.06</td>
<td>0.03</td>
</tr>
<tr>
<td>0.0994</td>
<td>6.15 ± 0.60</td>
<td>5.89 ± 0.59</td>
<td>0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>0.1002</td>
<td>3.67 ± 0.58</td>
<td>3.60 ± 0.57</td>
<td>0.80</td>
<td>0.04</td>
</tr>
<tr>
<td>0.1026</td>
<td>3.23 ± 0.56</td>
<td>3.23 ± 0.56</td>
<td>0.76</td>
<td>0.05</td>
</tr>
<tr>
<td>0.1034</td>
<td>1.86 ± 0.45</td>
<td>1.78 ± 0.44</td>
<td>0.48</td>
<td>0.12</td>
</tr>
<tr>
<td>0.1042</td>
<td>1.34 ± 0.15</td>
<td>1.28 ± 0.15</td>
<td>0.24</td>
<td>0.32</td>
</tr>
</tbody>
</table>

1.33 and 2.19 \(^{85,86}\) and with the simulation results from Okamoto and Hansmann \(^{39}\) which yield estimates of 1.1 to 1.7. The nucleation parameter, \( \sigma_{ZB} \) decreases with increasing temperature in the temperature range \( T^* = 0.08 \) to the transition temperature. This indicates that nucleation is harder to accomplish at the transition temperature than at lower temperatures. However, as temperature is increased above the transition temperature, \( \sigma_{ZB} \) increases, indicating that nucleation is easier at higher temperatures. Our values of \( \sigma_{ZB} \) are all much larger than the experimental estimates \( \approx 0.003 \) \(^{85,88,89}\), are in good agreement with the simulation estimates of Okamoto and Hansmann \( \approx 0.07 \) \(^{39,40}\), Sung and Wu \( \approx 0.056 \) \(^{90}\), García and Sanbonmatsu \( \approx 0.06 \) \(^{47}\), and Ohkuho and Brooks \( \approx 0.05 \) \(^{48}\), and are smaller than simulation estimates \( \approx 0.5 \) of Daggett and coworkers \(^{91,34}\).
3.3.2 Conformational transition as a function of temperature and hydrophobic interaction strength

We now consider how the strength of the hydrophobic interaction between non-polar sidechains affects the folding transition from the random coil to the native state. Figure 3.7 shows the percentage of $\alpha$-helical hydrogen bonds formed (relative to the perfect $\alpha$-helix value of 12) versus reduced temperature ($T^* = k_B T / \epsilon_{HB}$) at different ratios $R$ of the strength of the hydrophobic interaction ($\epsilon_{HP}$) relative to that of the hydrogen bond, $R \equiv \epsilon_{HP} / \epsilon_{HB}$. Notice that the reduced temperature $T^*$ not only measures temperature but also the strength of the hydrogen bond, ranging from strong hydrogen bonds at low $T^*$ (corresponding to deep $N$ and $C$ square wells) to weak bonds at high $T^*$ (corresponding to shallow $N$ and $C$ square wells). These results are average values calculated from a large ensemble of peptide configurations collected at each temperature during the replica-exchange simulations. Since the average errors, which are standard deviations from the average values at each temperature, are relatively small as those displayed in Figure 3.6, we plot only the average values, omitting the error bars. When the hydrophobic interaction is absent ($R=0$), which is equivalent to the vacuum condition, the midpoint of the folding transition (50% helicity) is at $T^*=0.100$, and the $\alpha$-helical native state is the dominant conformation in the reduced temperature range of 0.080 to 0.095. When the hydrophobic interaction is present (equivalent to the presence of solution) at a low value ($R=1/12$), the $\alpha$-helical native state region expands to higher temperatures as the transition temperature shifts to $T^*=0.106$. In this case, turning on the hydrophobicity results in having ten pairs of hydrophobic side chains interact in the native $\alpha$-helical structure, stabilizing an $\alpha$-helical state which would otherwise be unstable in the absence of hydrophobicity. Increasing the level of hydrophobicity
to $R=1/10$ further stabilizes the $\alpha$-helical by increasing the midpoint of the folding transition to $T^*=0.110$. Increasing the level of the hydrophobic interaction further to $R=1/8$, $R=1/6$, and then $R=1/4$ decreases the stability of the $\alpha$-helical state as the midpoints of these folding transitions are at $T^*=0.103$, 0.092, and 0.085 respectively. Interestingly, the stability of the $\alpha$-helical state at $R=1/8$ is almost the same as that for the system without hydrophobicity since their folding transition midpoints are nearly the same. Increasing the level of the hydrophobic interaction further to $R=1/2$ severely diminishes the peptide’s ability to form helices as indicated by the very low number of $\alpha$-helical hydrogen bonds seen in the figure. In summary, the results in Figure 3.7 indicate that small amounts of hydrophobicity stabilize the $\alpha$-helical native state of the model peptide but large amounts of hydrophobicity prevent $\alpha$-helix formation as the hydrophobic interactions overwhelm the system and trap the chain in non-native structures.

Increasing the hydrophobic interaction strength causes $\beta$ hydrogen bonds to form at the expense of the $\alpha$-helical hydrogen bonds. This is illustrated in Figure 3.8, which shows the number of $\beta$ hydrogen bonds formed versus reduced temperature $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction ($\epsilon_{HP}$) relative to that of the hydrogen bond ($\epsilon_{HB}$). At low hydrophobic interaction strengths ($R=0$, 1/12, and 1/10), the amplitude of each curve is less than five meaning that most $\beta$-structures formed at these hydrophobic interaction strengths are not $\beta$-hairpins. As the hydrophobic interaction strength increases further to $R=1/8$, 1/6 and then 1/4, the range of $\beta$-structure temperature region expands and the amplitudes of the curves increase, indicating that the chains are more likely to form more $\beta$ hydrogen bonds. The maximum number of $\beta$ hydrogen bonds at these $R$ values is well above the cut-off value for the $\beta$-hairpin structure meaning that most $\beta$-structures formed over the $\beta$-structure temperature re-
Region at these hydrophobic interaction strengths are $\beta$-hairpins. As the hydrophobic interaction strength increases to $R=1/2$, the trend with increasing $R$ changes in that the $\beta$ structures are no longer stable at intermediate temperatures but are instead stable at low temperatures. Besides $\beta$-hairpins, other $\beta$-structures that are $\beta$-sheet-like are formed at $R=1/2$.

The strength of the hydrophobic interaction between side chains affects the conformational transition between different states as can be seen in Figure 3.9, which shows the reduced potential energy of the system, $E^*$, versus reduced temperature $T^*$ at different ratios of the strength of the hydrophobic interaction ($\epsilon_{HP}$) relative to that of the hydrogen bond ($\epsilon_{HB}$). The reduced potential energy of the system, $E^*$, is the sum of the energy contributed by hydrogen bonds (the number of hydrogen bonds times $\epsilon_{HB}$) and the energy contributed by hydrophobic interactions (the number of hydrophobic interactions times $\epsilon_{HP}$) divided by $\epsilon_{HB}$. At $R < 1/2$, the $\alpha$-helical state is characterized by a reduced potential energy between -12.5 and -14.5 depending upon the value of $R$, the $\beta$-hairpin state is characterized by a reduced potential energy between -6.0 and -9.0, and the random coil state is characterized by a reduced potential energy near 0. As indicated by Figure 3.9, when the hydrophobic interaction strength is weak ($R=0, 1/12$ and $1/10$), there is a relatively sharp transition from the random coil ($E^*=0$) at high temperatures to the $\alpha$-helix at low temperatures. As the hydrophobic interaction strength increases up to $R=1/4$, there is a significant expansion in the range of temperatures over which $\beta$-hairpins are found. In this case, the usual $\alpha$-helix - random-coil transition is replaced by an $\alpha$-helix - $\beta$-hairpin - random-coil transition. As the hydrophobic interaction strength increases up to $R=1/2$, the potential energies of the $\alpha$-helical and $\beta$-structure states, which include $\beta$-hairpins and $\beta$-sheet-like structures, are about the same at -16.0. However, as already discussed earlier with Figures 3.7 and 3.8, instead of
forming an α-helix at low temperatures, the system forms β-hairpins and β-sheet-like structures (discussed below). In other words, as the hydrophobic interaction strength increases from $R=1/4$ to $R=1/2$, the potential energy of β-structure states decrease, moving towards that of the α-helical state. Thus, the transition at $R=1/2$ is between β-structure states at low temperatures and the random coil at high temperatures.

The conformational transitions found at different hydrophobic interaction strengths are displayed in Figure 3.10, which shows the probability of being in the different states as a function of the reduced temperature and (a) the number of α-helical hydrogen bonds ($n_\alpha$) and (b) the number of β hydrogen bonds ($n_\beta$) at $R=1/12$, $1/6$, and $1/2$. High values of the probability correspond to the color red whereas low values of the probability correspond to the color purple on the scale at the top of the figure. At $R=1/12$, Figure 3.10(a) indicates that there is a relatively broad region of low temperatures ($T^*=0.08-0.10$) over which the probability that $n_\alpha$ reaches its maximum value, $n_\alpha=12$ in the α-helical state, is high. (High probability corresponds to the color red on the scales at the top the figure.) Figure 3.10(b) indicates that there is a relatively narrow range of intermediate temperatures ($T^*=0.11-0.12$) over which the probability that $n_\beta$ reaches its maximum value, $n_\beta=7$ in the β-hairpin state, is high. When the hydrophobic interaction strength increases to $R=1/6$, Figure 3.10(a) indicates that the low temperature region over which $n_\alpha$ reaches its maximum value of $n_\alpha=12$ is considerably smaller ($T^*=0.08-0.09$) than that at $R=1/12$. In contrast, Figure 3.10(b) indicates that the intermediate temperature region over which $n_\beta$ reaches its maximum value of $n_\beta=7$ is significantly larger ($T^*=0.09-0.12$) than that at $R=1/12$. When the hydrophobic interaction strength increases further to $R=1/2$, the probability that $n_\alpha$ is sizable is negligible as seen in Figure 3.10(a) over the whole temperature range. Instead of forming
$\alpha$-helical hydrogen bonds at low temperatures, high numbers, $n_{\beta}=7$, of $\beta$ hydrogen bonds are formed as seen in Figure 3.10(b).

The free energy associated with the $\alpha$-helix - random-coil transition at low hydrophobic interaction strengths, $R=1/12$, has been calculated using the weighted histogram method. The result is given in Figure 3.11, which shows isosurface plots of: (a) the free energy versus the number of $\alpha$-helical ($n_\alpha$) and $\beta$ hydrogen ($n_\beta$) bonds, and (b) the free energy versus the total number of hydrogen bonds ($n_{HB}$) and the number of hydrophobic interactions ($n_{HP}$) at three temperatures $T^*=0.08$, 0.11 and 0.15. As previously seen in Figure 3.9, $T^*=0.08$ and 0.15 correspond to the lowest and highest temperatures at which the simulations at $R=1/12$ were conducted, and $T^*=0.11$ corresponds to the transition temperature. The region in Figure 3.11(a) where $n_\alpha \geq 8$ and $n_\beta \leq 1$ and the region in Figure 3.11(b) where $n_{HB} \geq 8$ and $8 \geq n_{HP} \geq 10$ belong to the $\alpha$-helical state. The region where $n_\alpha < 1$ and $n_\beta > 5$ in Figure 3.11(a) and the region where $5 < n_{HB} \leq 7$ and $8 \leq n_{HP} \leq 12$ in Figure 3.11(b) belong to the $\beta$-hairpin structure. The region where $n_\alpha < 3$ and $n_\beta < 3$ in Figure 3.11(a) and the region where $n_{HB} < 3$ and $n_{HP} < 4$ in Figure 3.11(b) belong to the random-coil state. At $T^*=0.08$, the $\alpha$-helical state has the lowest free energy, which corresponds to the color red on the scales at the right of the figure, the $\beta$-hairpin state has the second lowest free energy, and the random-coil state has the highest free energy. At $T^*=0.15$, the random coil has the lowest free energy. In other words, at the two extreme temperatures, the system exhibits two distinct states. At the intermediate temperature, the third state - the $\beta$-hairpin - does exist as seen in plots at $T^*=0.11$ but does not dominate in that it has a similar free energy to that in the other two states.

The free energy associated with the $\alpha$-helix - $\beta$-hairpin - random-coil transition at interme-
mediate hydrophobic interaction strengths, $R=1/6$, is shown in Figure 3.12, which depicts isosurface plots of the free energy versus (a) the number of $\alpha$-helical ($n_\alpha$) and the number of $\beta$ ($n_\beta$) hydrogen bonds, and (b) the total number of hydrogen bonds ($n_{HB}$) and the number of hydrophobic interactions ($n_{HP}$) for the system at $R=1/6$ at three temperatures $T^*=0.080, 0.106$ and $0.166$. Low free energy corresponds the color red on the scales at the right of each figure. As previously seen in Figure 3.9, $T^*=0.080$ and $T^*=0.166$ correspond to the lowest and highest temperatures at which the simulations at $R=1/6$ were conducted; $T^*=0.106$ corresponds to an intermediate temperature. As indicated by Figure 3.12, the $\alpha$-helical state is the most stable structure at $T^*=0.080$ while the random-coil state is the preferred structure at $T^*=0.166$. At $T^*=0.106$, the $\beta$-hairpin state has the lowest free energy as indicated by the red area in Figure 3.12. In other words, at the two extreme temperatures, the system exhibits two distinct states but at the intermediate temperature $T^*=0.106$, the third state - a $\beta$-hairpin - dominates.

The free energy associated with the $\beta$-structure states - random-coil transition at high hydrophobic interaction strengths, $R=1/2$, is shown in Figure 3.13, which depicts isosurface plots of the free energy versus (a) the number of $\alpha$-helical ($n_\alpha$) and the number of $\beta$ ($n_\beta$) hydrogen bonds, and (b) the total number of hydrogen bonds ($n_{HB}$) and the number of hydrophobic interactions ($n_{HP}$) for the system at $R=1/2$ at three temperatures $T^*=0.080, 0.185$ and $0.380$. At $T^*=0.080$, the polyalanine peptide is in multiple $\beta$-structure states with three to seven $\beta$ hydrogen bonds and more than thirteen hydrophobic interactions. Besides $\beta$-hairpins, other $\beta$-like structures such as those shown in Figure 3.5 were observed. These structures have more than one $\beta$-turn with less $\beta$ hydrogen bonds but more hydrophobic interactions than those found in a $\beta$-hairpin. Therefore, they are more compact than $\beta$-hairpins. At $T^*=0.185$, states with two
or less $\beta$ hydrogen bonds and 8 to 20 hydrophobic interactions are found; these are amorphous without any secondary structures. We consider this to be a molten globule state since it is neither fully folded nor fully unfolded. At $T^*=0.380$, the polyalanine peptide is a random coil with no $\beta$ hydrogen bonds and low numbers of hydrophobic interactions.

There are two factors that decide whether the polyalanine peptide will fold into an $\alpha$-helix or a $\beta$-structure: enthalpy and entropy. Although the $\alpha$-helix state has more intrapeptide hydrogen bonds than the $\beta$-structure states (the maximum is 12 compared to 7), it has a fixed number of hydrophobic interactions which are specific between certain residues whereas $\beta$-structures have a large ensemble of different states with different numbers of hydrophobic interactions which are relatively flexible. A measure of the number of states available to the system and hence the entropy is the size of the area occupied by these states in Figures 3.11, 3.12, and 3.13. For example, the area in Figures 3.11(b), 3.12(b), and 3.13(b) occupied by the $\beta$-structure states is noticeably larger than those occupied by other states, which means that the entropy for the $\beta$-structure states is relatively large. As the hydrophobicity is increased, the enthalpies of both the $\alpha$-helix and $\beta$-structure states decrease; however, the decrease in the enthalpy of the $\beta$-structure states is more significant than that of the $\alpha$-helical state since the $\beta$-structure states can increase their number of hydrophobic interactions. In fact, the enthalpy of the $\beta$-structure states decreases towards that of the $\alpha$-helical state as $R$ increases (see Figure 3.9). At low to intermediate hydrophobic interaction strengths ($R < 1/2$) and low temperature, the $\alpha$-helical state has the lowest enthalpy which, since $T^*$ is low, is the most dominant contribution to the free energy; therefore, it is stable, having the lowest free energy at low temperatures. However, as the temperature increases towards intermediate temperatures, the entropy starts to play an
important role in lowering the free energy and thus stabilizing the \( \beta \)-structure states, especially the \( \beta \)-hairpin state. As the hydrophobic interaction strength increases to \( R=1/2 \), the enthalpy of the \( \beta \)-structure states decreases below that of the \( \alpha \)-helical state, causing the \( \beta \)-structure states to have the lowest enthalpy which, coupled with their high entropy, means that the \( \beta \)-structure states have the lowest free energy over a wide range of temperatures (both intermediate and low). In other words, increasing the hydrophobicity stabilizes the \( \beta \)-structure states and thus destabilizes the \( \alpha \)-helical state.

### 3.4 Summary and Discussion

In this work, we presented a study of the folding thermodynamics of a model polyalanine peptide using an improved version of our previously developed intermediate-resolution protein model in conjunction with discontinuous molecular dynamics simulations. The focus was on how the solvent condition, as measured by the strength of the hydrophobic interaction between non-polar sidechains and temperature, affects the peptide’s conformational transitions. We observe that the environment has a significant effect on the stability of the various structures, \( \alpha \)-helix, \( \beta \)-structures (including \( \beta \)-hairpins and \( \beta \)-sheet-like structures) and random coil, at different temperatures. At low hydrophobic interaction strength, the peptide exhibits a two-state folding transition in which the \( \alpha \)-helical state is the most stable structure at low temperatures and the random-coil state is the most stable state at high temperatures. At intermediate hydrophobic interaction strength, the \( \beta \)-hairpin state is introduced, resulting in an \( \alpha \)-helical conformation at low temperatures, a \( \beta \)-hairpin at intermediate temperatures, and a random-coil at high temperatures. At very high values of the hydrophobic interaction strength,
the peptide folds via a two-state mechanism in which different types of \( \beta \)-structures including \( \beta \)-hairpin and \( \beta \)-sheet-like structures are the most stable structures at low temperatures and the random-coil state is the most stable state at high temperatures. Although there is no experimental evidence for the \( \alpha \)-\( \beta \) transition in polyalanines as temperature changes, there are experiments which show that polyalanine adopts an \( \alpha \)-helical conformation in hydrophobic environments such as the solid state or in non-polar organic solutions and a \( \beta \)-structure conformation in polar aqueous solution \(^{93,89,94–96,60,97}\). This is also observed in experiments on many heterogeneous peptides which can be folded into alternative stable structures by changing the solution conditions such as the pH, salt or organic cosolvent concentration, peptide concentration, and the redox state \(^{98,3–6,14,7,8,16,9,10,17,12,13,11,15}\). Although there are many computer simulation studies in which the transition between \( \alpha \)-helix and random coil of polyalanines is observed \(^{39–41,43,42,44,49,51,45–48}\), there are only two studies thus far that predict the formation of both \( \alpha \)-helix and \( \beta \)-structures \(^{52,53}\). The vast majority of simulation studies by other investigators mentioned here are based on all-atom models, which prevents them from exploring transitions between different peptide structures over a wide range of solvent conditions and temperatures. Recently, simulations by Vila \textit{et al.}\(^{37}\), Hansmann, Okamoto \textit{et al.}\(^{43–45}\), and Olivella \textit{et al.}\(^{49}\) on polyalanines have been conducted over a wide range of temperatures but only in the presence and absence of the solvent. They showed that intra-helical hydrogen bonds are weakened in the presence of polar solvent, which might suggest that other non-helical interactions such as \( \beta \) hydrogen bonds are preferred.

It is important to point out that our model and analysis are subject to a number of limitations. Though our model’s backbone is detailed enough to display genuine protein character,
the model side chains are very simplified. Our model can account for the hydrophobic force by allowing side chain beads to be either hydrophobic or polar but ignores other aspects of side chain character including charge. Another shortcoming of the model is the lack of solvent. We incorporate solvent as a potential of mean force, which means that all chain beads are equally affected by “solvent” regardless of the chain conformation. Consequently beads in the interior of a compact structure are just as affected by solvent as they would be if they were fully exposed in a random coil structure. A more accurate solvation model would allow forces such as the hydrogen bonding force in the core of a collapsed chain to be different from those at the surface. It is unclear, however, that this shortcoming is critical in the work presented here since our peptides are so small that the concept of a “core” is hard to define. Despite these limitations, our model allows us to conduct simulations over a wide range of solvent conditions and temperatures because of its speed. Ongoing research in our group is aimed at using this simple protein model to study the effects of varying the temperature, peptide concentration and hydrophobic interaction strength on the formation of ordered aggregates, e.g. fibrils, in multi-protein environments.

3.5 Acknowledgments

The authors are grateful to Robert Baldwin, Sylvie Blondelle, Ken Dill, Angel García, Jeffery Kelly, Dan Kirschner, Hugh Nymeyer, and Sheena Radford for helpful discussions. They are also grateful to Andrew Schultz for helping us to apply his efficiency techniques to our simulation code. This work was supported by the National Institutes of Health under grant number GM-56766 and the National Science Foundation under grant number CTS-9704044.
3.6 References


3.7 Figures

3.1 Geometry of the intermediate resolution protein model for polyalanine. Covalent bonds are shown with narrow solid lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_\alpha$ distances, and residue L-isomerization. Note that the united atoms are not shown full size for ease of viewing. .......................... 108

3.2 Ramachandran plots for simulations of alanine residues based on the bond length and pseudo-bond tolerance at (a) 2.0% (taken from our previous study) and (b) 2.375%. .......................... 109

3.3 Backbone hydrogen bonding. ................................................. 110

3.4 Distribution of hydrogen bond (a) angles and (b) distances for $\alpha$-helices, and (c) angles and (d) distances for $\beta$-structures. ............................................. 111

3.5 Snapshots of the 16-residue peptide in (a) an $\alpha$-helix and (b) a $\beta$-hairpin. Besides $\beta$-hairpins, other $\beta$-structures are $\beta$-sheet-like with (c) two $\beta$-turns, and (d) three $\beta$-turns. Hydrophobic sidechains are white, polar sidechains (at chain ends) are dark gray, and backbone atoms are light gray. Note that the united atoms are not shown full size for ease of viewing. ........................................... 112

3.6 Percent $\alpha$-helical hydrogen bonds formed as a function of $T^* = k_BT/\epsilon_{HB}$ in the absence of hydrophobicity. ........................................... 113
3.7 Percent $\alpha$-helical hydrogen bonds formed as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$. .......................... 114

3.8 Number of $\beta$-structure hydrogen bonds formed as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$. .......................... 115

3.9 Reduced potential energy of the system, $E^* = E/\epsilon_{HB}$ as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$: $R=0$ (filled circles), $R=1/12$ (open circles), $R=1/10$ (filled squares), $R=1/8$ (open squares), $R=1/6$ (filled diamonds), $R=1/4$ (open diamonds), and $R=1/2$ (filled triangles). .......................... 116

3.10 The conformational probability as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and $T^*$ and (b) the number of $\beta$ hydrogen bonds $n_\beta$ and $T^*$ at $R = 1/12$, 1/6, and 1/2. .......................... 117

3.11 The Helmholtz free energy of the system at $R=1/12$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures. .......................... 118

3.12 The Helmholtz free energy of the system at $R=1/6$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures. .......................... 119
3.13 The Helmholtz free energy of the system at $R=1/2$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures.
Figure 3.1: Geometry of the intermediate resolution protein model for polyalanine. Covalent bonds are shown with narrow solid lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_\alpha$ distances, and residue L-isomerization. Note that the united atoms are not shown full size for ease of viewing.
Figure 3.2: Ramachandran plots for simulations of alanine residues based on the bond length and pseudo-bond tolerance at (a) 2.0% (taken from our previous study)\textsuperscript{54} and (b) 2.375%.
Figure 3.3: Backbone hydrogen bonding is achieved through a square-well interaction between \( N_i \) and \( C_j \). The dashed circles symbolize the square-well attractive shells around \( N_i \) and \( C_j \). The location of \( H_i \) can be determined by examining the geometry of \( C_{i-1}, N_i, \) and \( C_{\alpha,i} \) and then projecting down through \( N_i \) as shown by the arrow. The location of \( O_j \) can be determined by examining the geometry of \( C_{\alpha,j}, C_j, \) and \( N_{j+1} \) and then projecting up through \( C_j \) as shown by the arrow. The thick disjointed lines symbolize the square-shoulder interactions between auxiliary pairs \( N_i-C_{\alpha,j}, N_i-N_{j+1}, C_j-C_{\alpha,i}, \) and \( C_j-C_{i-1}. \)
Figure 3.4: Distribution of hydrogen bond (a) angles and (b) distances for $\alpha$-helices, and (c) angles and (d) distances for $\beta$-structures.
Figure 3.5: Snapshots of the 16-residue peptide in (a) an α-helix and (b) a β-hairpin. Besides β-hairpins, other β-structures are β-sheet-like with (c) two β-turns, and (d) three β-turns. Hydrophobic sidechains are white, polar sidechains (at chain ends) are dark gray, and backbone atoms are light gray. Note that the united atoms are not shown full size for ease of viewing.
Figure 3.6: Percent $\alpha$-helical hydrogen bonds formed as a function of $T^* = k_B T/\epsilon_{HB}$ in the absence of hydrophobicity.
Figure 3.7: Percent $\alpha$-helical hydrogen bonds formed as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$. 
Figure 3.8: Number of $\beta$-structure hydrogen bonds formed as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$. 
Figure 3.9: Reduced potential energy of the system, $E^* = E/\epsilon_{HB}$ as a function of $T^*$ at different ratios $R$ of the strength of the hydrophobic interaction, $\epsilon_{HP}$, relative to that of the hydrogen bond, $\epsilon_{HB}$: $R=0$ (filled circles), $R=1/12$ (open circles), $R=1/10$ (filled squares), $R=1/8$ (open squares), $R=1/6$ (filled diamonds), $R=1/4$ (open diamonds), and $R=1/2$ (filled triangles).
Figure 3.10: The conformational probability as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and $T^*$ and (b) the number of $\beta$ hydrogen bonds $n_\beta$ and $T^*$ at $R^* = 1/12$, 1/6, and 1/2.
Figure 3.11: The Helmholtz free energy of the system at $R=1/12$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures.
Figure 3.12: The Helmholtz free energy of the system at $R=1/6$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures.
Figure 3.13: The Helmholtz free energy of the system at $R=1/2$ as a function of (a) the number of $\alpha$-helical hydrogen bonds $n_\alpha$ and the number of $\beta$ hydrogen bonds $n_\beta$ and (b) the total number of hydrogen bonds $n_{HB}$ and the number of hydrophobic interactions $n_{HP}$ at three temperatures. 

(a) $n_\alpha$ vs. $n_\beta$

(b) $n_{HB}$ vs. $n_{HP}$
CHAPTER 4

MOLECULAR DYNAMICS SIMULATIONS
OF SPONTANEOUS FIBRIL FORMATION
FROM RANDOM-COIL PEPTIDES

4.1 Introduction

The assembly of normally soluble proteins into ordered aggregates, known as amyloid fibrils, is a cause or associated symptom of numerous human disorders, including Alzheimer’s, Parkinson’s and Huntington’s diseases, the prion diseases and adult-onset diabetes \(^{1–5}\). In each of these disorders, which are known collectively as the amyloidoses, a specific protein slowly accumulates in fibrillar tangles or plaques, destroying the architecture and function of the surrounding tissue, usually with degenerative and ultimately-fatal consequences. Although little is known about the molecular basis for fibrillization, tantalizing clues emerge when the common features of the various amyloidoses are examined. Fibrils of amyloidogenic proteins formed \textit{in vitro} exhibit strikingly similar morphologies despite a lack of similarity in their sequence, structure and function \(^{6–9,4}\). They are invariably long, straight and unbranched, and consist of two or more smaller fibrils, called protofilaments (and sometimes protofibrils) which are themselves
long ribbons of layered crossed β-sheets propagating along the fibril axis. Recent evidence that proteins other than those associated with amyloid diseases form fibrils in vitro under mildly denaturing conditions has led leaders in the field to suggest that fibril formation is an intrinsic property of polypeptides, albeit under appropriate conditions. This implies that the forces that stabilize fibrils are the forces common to all proteins - hydrophobic interactions and backbone hydrogen bonding - and not the forces associated with specific interactions between sidechains. It follows then that progress towards understanding the origins of various protein deposition diseases can be made by in vitro examination of the general features of protein fibrillization using model proteins that are less complex than the specific amyloidogenic protein. Progress could also be made by in silico examination of the molecular-level mechanisms responsible for the formation of fibrils by model peptides so long as the protein geometry, hydrophobicity and hydrogen bonding are properly taken into account.

Most computer simulations of fibril-forming peptides have been devoted to the study of isolated peptides and have employed high-resolution protein models, which are based on a realistic representation of protein geometry and a fairly faithful accounting for the energetics of every atom on the protein and on the solvent. In the past few years, however, there have been a number of attempts to simulate multi-peptide systems containing already-formed amyloid fibrils. Li et al. performed molecular dynamics (MD) simulations in explicit aqueous solution using the Amber 4.1 package with the Cornell et al. forcefield on Aβ amyloid protofilaments to test the stability of proposed fibril structures. George et al. conducted MD simulations on β-amyloid using the CHARMM package to determine the difference between inter-peptide interactions in Aβ(1-40) and Aβ(1-43) and between the native and mutated forms
of β-amyloid. They also examined the interaction between the surface of a βA(1-43) fibril and a proposed aggregation-inhibitor, the molecule IDOX. Nussinov et al. have conducted MD simulations using the packages Discover and CHARMM to study the stability of fibrils containing model peptides arranged in different parallel/antiparallel stacking arrangements. The model peptide systems studied include AGAAAAGA and AAAAAAAA; fragments of the Alzheimer’s β-amyloid peptide, Aβ16–22, Aβ16–35, and Aβ10–35; and segments of the human islet amyloid polypeptide 22-27 (NFGAIL) and 22-29 (NFGAILSS). Lakdawala et al. performed MD simulations using the AMBER package to study the stability of amyloid fibrils containing fragments of two to six residues of the hydrophobic core sequence KLVFFA of Aβ(10-35). Hwang et al. conducted MD simulations using the CHARMM package on the supramolecular structure of helical ribbons containing the β-sheet peptide KFE8 to see which packing geometries yielded the most stable structure. Kuwata et al. performed MD simulations using DISCOVER 2.98 to study the stability of fibrils containing eight mouse prion protein fragments 106-126 in different parallel/antiparallel stacking arrangements.

High-resolution simulation studies of the formation of fibrils from random coils have been conducted but they are generally limited to early events in the assembly of a few peptides. Gsponer et al. performed MD simulations on a β-sheet of the heptapeptide GNNQQQNY from the yeast prion Sup35 using the CHARMM program and observed the formation of a β-sheet from a system containing three peptides. Mager et al. conducted MD simulations on a system containing two to four Aβ1–42 peptides using the AMBER program and observed the formation of a β-sheet. Fernandez et al. performed MD simulations on systems containing two Aβ(12-24) or two β2-microglobulin(21-31) fragments using the AMBER95 forcefield to
study how crowding affects aggregation. Klimov et al.\textsuperscript{44} conducted long MD simulations on systems containing three A\textsubscript{16–22} peptides using the MOIL program \textsuperscript{45}. Although the above-mentioned studies offer considerable insight into the properties of fibrils and the mechanisms by which they are formed, the systems considered do not contain enough peptides to mimic the nucleus that stabilizes the large fibrils observed in experiments. An alternative approach was taken by Ma et al.\textsuperscript{32} who conducted MD simulations using the package Discover 2.98 for up to 4 nanoseconds on a system containing an already-formed fibrillar aggregate of eight AAAAAAAA peptides that is surrounded by either a single α-helical AAAAAAAA monomer or eight random-coil AAAAAAAA peptides. They did not observe fibril growth due to the limited simulation time in either case. Given current computational capabilities, simpler models are required to simulate multi-protein systems.

Computer simulations using low-resolution models, which are based on a coarse-grained representation of protein geometry and energetics, have been used by a few investigators to study fibril formation. The simplest low-resolution protein representation, of course, is the HP lattice model, which captures the hydrophobic effect by representing each bead on the chain as a hydrophobic (H) or polar (P) residue \textsuperscript{46}. Harrison et al.\textsuperscript{47} examined pairs of two- and three-dimensional HP lattice chains to learn about the thermodynamics of the conformational change associated with aggregation of model prion proteins. There have been several other Monte Carlo or exact enumeration studies of the aggregation of a few lattice model peptides \textsuperscript{48–55} but these system sizes are too small to provide a full exploration of the competition between folding and aggregation. In contrast, the lattice systems simulated by Gupta and Hall \textsuperscript{56,57} and Nguyen and Hall \textsuperscript{58} were truly multichain systems, containing between 20 and 50 protein chains. Although
the lattice model studies described here provide invaluable insights into the basic physics underlying protein aggregation in general, they do not adequately account for the different forces, such as hydrogen bonding, that play an important role in fibril formation.

Intermediate-resolution protein models \(^{59-70}\), which are essentially a compromise between the simplified chain models described above and the detailed all-atom models like AMBER \(^{71}\), CHARMM \(^{72}\), and ENCAD \(^{73}\), have been used extensively in recent years to simulate the folding of isolated proteins. They have also been used by a few investigators to study fibril formation. Jang et al.\(^{74,75}\) applied an off-lattice model with each amino acid residue represented by a single bead interacting via Go potentials (which have a partial built-in bias towards the native state) to study the thermodynamics and kinetics of the assembly of four model \(\beta\)-sheet peptides into a tetrameric \(\beta\)-sheet complex. Ding et al.\(^{76}\) applied an off-lattice model with each amino acid residue represented by one backbone bead and one sidechain bead interacting via Go potentials to study the formation of a fibrillar double \(\beta\)-sheet structure containing eight model Src SH3 domain proteins. These simulations provide more detail on the mechanisms that govern fibril formation than the lattice models. However, since the Go potential contains a built-in bias toward the native conformation, they are not suitable for the study of spontaneous fibril formation from random-coil configurations.

An intermediate-resolution protein model that has no built-in bias toward any conformation has been developed by Smith and Hall \(^{70,77,78}\) and later improved by Nguyen and Hall \(^{79}\). Each amino acid is represented via an united atom approach - three beads for the backbone and one bead for the sidechain. The model contains enough genuine protein character to mimic real protein dynamics yet is simple enough to be computationally tractable, especially for use in
computer simulations of protein aggregation in relatively large systems. This model is designed
to be used with discontinuous molecular dynamics (DMD)\textsuperscript{80–83}, an extremely fast alternative
to traditional molecular dynamics, that is applicable to systems of molecules interacting via
discontinuous potentials, e.g., hard-sphere and square-well potentials. The solvent is modeled
implicitly by including the hydrophobic interaction between non-polar sidechains. Backbone
hydrogen bonding is modeled in explicit detail. Using this DMD algorithm, we are able to
sample much wider regions of conformational space, longer time scales, and larger systems than
in traditional molecular dynamics. As we will show in this paper, we are able to simulate the
formation of fibrils from systems containing between 12 and 96 16-residue polyalanine peptides
starting from the random state.

The model peptide chosen for study is the polyalanine-based peptide Ac-KA\textsubscript{14}K-NH\textsubscript{2}. We
focus on polyalanine-based peptides for three reasons. First, the small, uncharged, unbranched
nature of alanine residues is amenable to simulation with the intermediate-resolution protein
model that we developed previously\textsuperscript{70,77}. Second, polyalanine repeats have been implicated in
human pathologies, notably in the formation of anomalous filamentous intranuclear inclusions
in oculopharyngeal muscular dystrophy patients\textsuperscript{84}. Third, synthetic polyalanine-based peptides
have been shown by Blondelle \textit{et al.} to undergo a transition from \(\alpha\)-helical structures to \(\beta\)-sheet
complexes \textit{in vitro}\textsuperscript{85,86}, mimicking the structural transition that is believed to be a prerequi-
site for fibril nucleation and growth\textsuperscript{87–89,690,47,91}. Blondelle \textit{et al.} observed that the \(\alpha\)-helical
structures were stabilized in part by intramolecular \(\alpha\)-helical bonds and the macromolecular \(\beta\)-
sheet complex was stabilized by hydrophobic inter-sheet interactions. Using circular dichroism,
fourier-transform infrared spectroscopy, and reversed phase high performance liquid chromatog-
raphy, they found that: (1) \(\beta\)-sheet complex formation increased with increasing temperature, exhibiting an S-shaped dependence on temperature with a critical temperature of 45°C at a peptide concentration of 1.8mM and an incubation time of 3h, and (2) \(\beta\)-sheet complex formation increased with increasing peptide concentration above a critical concentration of 1mM at 65°C.

In this paper, we investigate how peptide concentration and temperature affect the formation of various Ac-K\(\alpha\)14K-NH\(_2\) structures including \(\alpha\)-helices, \(\beta\)-sheets and fibrils. Simulations are conducted on systems of 12, 24, 48, and 96 model 16-residue peptides at a wide variety of concentrations and temperatures by applying the discontinuous molecular dynamics simulation algorithm to our intermediate-resolution protein model. All simulations are performed in the canonical ensemble starting from a random coil configuration equilibrated at a high temperature and then slowly cooled to the temperature of interest so as to minimize kinetic trapping in local free energy minima. The percentage of peptides that form \(\alpha\)-helices, \(\beta\)-sheets or fibrils is monitored during the simulation. In addition, structural characteristics such as the peptide arrangement and packing of fibrils are examined and compared with those observed in experiments. We also study the overall stability of fibrils by conducting simulations on already-formed fibrils over a wide range of temperatures to investigate the relative importance of hydrogen bonding and hydrophobic interactions on fibril stability. The stability of our fibrillar structures is evaluated by comparing the abilities of the system to maintain the fibrillar structures at various temperatures which are higher than the fibril formation temperature.

Highlights of our results are the following. We observe the formation of fibrils from random coils in a relatively short period of time ranging between 40 to 160 hours on a single processor of an AMD Athlon MP 2200+ workstation for the 48-peptide system. To our knowledge, these are
the first simulations to span the whole process of fibril formation from the random coil state to
the fibril state on such a large system. We find that at low concentrations, random-coil peptides
assemble into α-helices at low temperatures and random coils at high temperatures. At inter-
mediate concentrations, random-coil peptides assemble into α-helices at low temperatures and
large β-sheet structures at high temperatures. At high concentrations, random-coil peptides form
β-sheets over a wide range of temperatures. These β-sheets assemble into fibrils above a critical
temperature that decreases with concentration and exceeds the isolated peptide’s folding tem-
perature. At very high temperatures and all concentrations, the system is in a random-coil state.
These results are in good qualitative agreement with those by Blondelle et al. on Ac-KA_{14}K-NH_{2}
peptides.\textsuperscript{85,86} Fibril formation appears to be a nucleated conformational conversion process in
that the formation of small amorphous aggregates precedes the formation of β-sheets and then
ordered nucleus and the subsequent rapid growth of a stable fibril. Once the nucleus has been
formed, fibril growth involves both elongation, in which monomeric peptides are added to the
end of the fibril, and lateral growth, in which already-formed β-sheets are added to the side of
the fibril. The fibrils observed in our simulations mimic the structural characteristics observed in
experiments in that most peptides in our fibrils are arranged in an in-register parallel orientation
with intra-sheet and inter-sheet distances that are similar to those observed in experiments, and
contain about six β-sheets each of which contains multiple peptides. Finally we find that when
the strength of the hydrophobic interaction between non-polar sidechains relative to the strength
of hydrogen bonding is high, the system forms amorphous rather than fibrillar aggregates.

This paper is organized as follows. Section 2 describes the protein model used for this work,
including the physical representation and the potential energy function, and the DMD simulation
technique. Section 3 presents and discusses the results obtained from: (1) simulations starting from random coils, and (2) simulations starting from already-formed fibrils. Section 4 provides a summary and a brief discussion of this work.

4.2 Method

4.2.1 Model peptide and forces

The model peptide has the sequence PH$_{14}$P, where H is a hydrophobic amino acid residue and P is a polar amino acid residue. This sequence was chosen to mimic Ac-KA$_{14}$K-NH$_2$ peptides which have been shown to form stable, soluble $\beta$-sheet complexes. The peptide is modeled at an intermediate level of resolution based on an united-atom approach with realistic bond lengths and bond-angle constraints that has the ability to interact both intra- and inter-molecularly via hydrogen bonding and hydrophobic interaction potentials. The geometry of the protein model is illustrated in Figure 4.1. Each amino acid residue is composed of four spheres, a three-sphere backbone comprised of united atom NH, C$_\alpha$H, and C=O, and a single-bead sidechain R (labeled N, C$_\alpha$, C, and R, respectively in the figure). All backbone bond lengths and bond angles are fixed at their ideal values; the distance between consecutive C$_\alpha$ atoms is fixed so as to maintain the inter-peptide bond in the trans configuration. The sidechains are held in positions relative to the backbone so that all residues are L-isomers. Details of the model including values for all parameters are given in our earlier papers.

The solvent is modeled implicitly in the sense that its effect is factored into the energy function as a potential of mean forces. All forces are modeled by either hard-sphere or square-well potentials. The excluded volumes of the four united atoms are modeled using hard-sphere
potentials with realistic diameters. Covalent bonds are maintained between adjacent spheres along the backbone by imposing hard sphere repulsions whenever the bond lengths attempt to move outside of the range between \( l(1-\delta) \) and \( l(1+\delta) \) where \( l \) is the bond length and \( \delta \) is a tolerance which we set equal to 2.375%. Ideal backbone bond angles, \( C_\alpha-C_\alpha \) distances and residue L-isomerization are achieved by imposing pseudobonds, as shown in Figure 4.1, which also fluctuate within a tolerance of 2.375%. Interactions between hydrophobic sidechains are represented by a square-well potential of depth \( \epsilon_{HP} \) and range \( 1.5\sigma_R \) where \( \sigma_R \) is the sidechain diameter. Hydrophobic sidechains must be separated by at least three intervening residues in order to interact. Hydrogen bonding between amide hydrogen atoms and carbonyl oxygen atoms on the same or neighboring chains are represented by a square-well attraction between NH and C=O united atoms whenever: (1) the virtual hydrogen and oxygen atoms (whose location can be calculated at any time) are separated by 4.2\( \AA \) (the sum of the NH and C=O well widths), (2) the nitrogen-hydrogen and carbon-oxygen vectors point towards each other within a fairly generous tolerance, (3) neither the NH nor the C=O are already involved in a hydrogen bond with a different partner, and (4) the NH and C=O are separated by at least three intervening residues along the chain. To satisfy the second requirement, the separations between the four auxiliary pairs, \( N_i-C_{\alpha,j} \), \( N_i-N_{j+1} \), \( C_{j-1}-C_{\alpha,i} \), \( C_{j-1}-C_{i-1} \), surrounding the hydrogen bond in question are limited to certain distances that are chosen to maintain ideal hydrogen bond angles. This is accomplished by imposing square-shoulder interactions between the auxiliary pairs. Besides adding stability to the hydrogen bond, these interactions exact a penalty for breaking a hydrogen bond when any one of these auxiliary pairs moves inside the specified separation and thus distorts the hydrogen bond angle. For more details on the hydrogen bonding model used here, see a
recent paper by Nguyen and Hall. For simplicity, the strength of a hydrophobic contact, $\epsilon_{HP}$, is fixed at 1/10, 1/8 and 1/6 the strength of a hydrogen bond, $\epsilon_{HB}$. Hydrogen bond strength and hydrophobic contact strength are independent of temperature, as has been assumed in previous simulation studies.

### 4.2.2 Discontinuous molecular dynamics

Simulations are performed using the discontinuous molecular dynamics (DMD) simulation algorithm, which is an extremely fast alternative to traditional molecular dynamics and is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials. Unlike soft potentials such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide, enabling the exact (as opposed to numerical) solution of the collision dynamics. DMD simulations proceed in the following fashion. The initial positions of the united atoms on the model protein are assigned at random, but may not violate any of the size constraints or assigned bond lengths and angles. The initial velocities are chosen at random from a Maxwell-Boltzmann distribution at a fixed reduced temperature, $T^* = k_B T / \epsilon_{HB}$, where $k_B$ is Boltzmann’s constant, $T$ is the temperature, and $\epsilon_{HB}$ is the depth of the square-well hydrogen bond potential. The simulation proceeds according to the following schedule: identify the first event, move forward in time until that event occurs, calculate new velocities for the pair of united atoms involved in the event and calculate any changes in system energy resulting from hydrogen bond events or hydrophobic interactions, find the second event, and so on. Types of events include excluded volume events, bond events, and square-well hydrogen bond and hydrophobic interaction events. An excluded volume event occurs when the
surfaces of two hard-sphere united atoms collide and repel each other. A bond (or pseudobond) event occurs via a hard-sphere repulsion when two adjacent spheres attempt to move outside of their bond length. Square well events include well-capture, well-bounce, and well-dissociation “collisions” when a sphere enters, attempts to leave, or leaves the square well of another sphere. For more details on DMD simulations with square-well potentials, see papers by Alder and Wainwright \textsuperscript{80} and Smith, Hall, and Freeman \textsuperscript{93}.

Simulations are performed in the canonical ensemble which means that the number of particles, volume, and temperature are held constant. Constant number of particles and volume are achieved by creating a virtual three-dimensional box for the simulation and allowing the model protein chains to move within that box. Periodic boundary conditions are used to eliminate artifacts due to simulation box wall effects. The dimensions of the box are chosen to ensure that a chain cannot interact with more than one image of any other chain. Constant temperature is achieved by implementing the Andersen thermostat method \textsuperscript{94} as was used previously \textsuperscript{70,95}. With this procedure, all united atoms in the simulation are subject to random collisions with ghost particles. The post-event velocity of an united atom colliding with a ghost particle is chosen randomly from a Maxwell-Boltzmann distribution at the simulation temperature.

For the fibril formation simulations, each system was started from a random coil configuration equilibrated at a high temperature, $T^* = 0.16$ and then cooled down to the temperature of interest ($T^* = 0.08$, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14, and 0.15) to minimize kinetic trapping in local free energy minima. The cooling rate is set by varying the number of ghost particles; the more ghost particles there are, the quicker the system reaches the temperature of interest. An initial test run at different cooling rates demonstrates that if the system is quickly quenched to
a low temperature, all peptides rapidly assemble into an amorphous aggregate whose energy is higher than that of the fibril obtained by cooling the system down slowly. Since we are interested in observing fibrils obtained at equilibrium, all systems were cooled down slowly; the number of ghost collisions was set at 0.005% of the total number of collisions during a simulation. The resulting cooling rate is $\Delta T^*/\Delta t^*=0.0004$, where $t^*$ is the reduced time which is defined as $t/\sigma\sqrt{k_B T/m}$ with $t$ the simulation time, and $\sigma$ and $m$ the average united atom diameter and mass. For these simulations, we use cubic boxes with sides ranging from 158 to 542 Å in length depending on the peptide concentration.

Although simulations were conducted on systems containing 12, 24, 48, and 96 peptides, we focus most of our analysis on the 48-peptide system. The peptide packing fraction for the 48-peptide system ranges from 0.000026 to 0.001056; here we have approximated the peptide volume to be $V_p \equiv 4\pi R_g^3/3$ where $R_g=9.98$ Å is the average peptide radius of gyration in our simulations when the peptide is in the random-coil conformation. The peptide concentrations for the 48-peptide system are $c \equiv N/N_A V \approx 0.5, 1, 2.5, 5, 10,$ and $20$mM where $N=48$ is the number of peptides, $N_A$ is Avogadro’s number, and $V$ the total system volume. Though our system is small compared with real fibrils which tend to contain four to six $\beta$-sheets with 1000 or more peptides per sheet $^{11,96–98}$, it is quite large compared with the few peptides simulated in other studies $^{39–43}$. Hopefully this is sufficient to provide a foundation for the basic understanding of fibril formation in larger systems. Simulations were performed for between 8 and 32 billion events, which required 40 hours at $T^*=0.08$ for 8 billion events to 160 hours at $T^*=0.15$ for 32 billion events on a single processor of an AMD Athlon MP 2200+ workstation for the 48-peptide system. All systems were simulated for long periods of time until the ensemble averages of the
system’s total potential energy varied by no more than 2.5% during the last three quarters of each simulation run.

Our results are reported in terms of the average percentage of peptides in the system that form the structures of interest such as \(\alpha\)-helices, aggregates, \(\beta\)-sheets and fibrils, which are defined in the following way. If 12 intra-peptide \(\alpha\)-helical hydrogen bonds are formed, the structure is an \(\alpha\)-helix. If each peptide in a group of peptides has at least two inter-peptide hydrogen bonds or hydrophobic interactions with a neighboring peptide in the same group, then that group is classified as an aggregate. Aggregates can be either ordered or amorphous. If an aggregate contains \(\beta\)-sheets or better yet fibrils, we classify it as an ordered aggregate; otherwise, we classify it as an amorphous aggregate. If each peptide in a group of peptides has at least seven inter-peptide \(\beta\) hydrogen bonds to a particular neighboring peptide in the group, we classify this group as a \(\beta\)-sheet. If at least two \(\beta\)-sheet structures form inter-sheet hydrophobic interactions (at least four hydrophobic interaction per peptide per \(\beta\)-sheet) and the \(\beta\)-sheet structures are at an angle that is less than 35\(^\circ\), we classify this as a fibril.

For the fibril stability simulations, each system was started from a 48-peptide fibrillar structure, which was formed at \(c=5\text{mM}\) and \(T^*=0.11\) by the slow-cooling method described above, and then quickly heated to the temperature of interest. The number of ghost collisions was set at 0.05\% of the total number of collisions during a simulation, which is sufficient to maintain the system temperature at a constant desired value. These simulations were performed at seven different reduced temperatures \(T^* = 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, \) and \(0.18\), each for 10 billion events, which required 50 hours on a single processor of an AMD Athlon MP 2200+ workstation.
4.3 Results and Discussion

Since this paper builds upon previous work by Nguyen and Hall on the folding thermodynamics of a single peptide of the same sequence, it is useful to briefly review those results that are pertinent to the discussion here. The peptide’s thermodynamics was explored using the replica exchange simulation method to map out the conformational transitions between the random coil, α-helix and β-structure at different temperatures. The effect of solvent conditions on the peptide’s stability was also investigated by varying the ratio $R$ of the strength of the hydrophobic interaction ($\epsilon_{HP}$) to that of the hydrogen bond ($\epsilon_{HB}$), $R \equiv \epsilon_{HP}/\epsilon_{HB}$. At each hydrophobic interaction strength, we calculated the conformational free energy of each structural state to determine the structure with the lowest free energy at each temperature and identify the conformational transition between different structural states at different temperatures. We find that peptides in our simulations tend to mimic real polyalanines in that they can exist in three distinct structural states: α-helices, β-structures, and random coils, depending upon the solvent conditions. At low values of the hydrophobic interaction strength between non-polar sidechains, i.e. $R = 0, 1/12$ and $1/10$, the polyalanines undergo a relatively sharp transition between an α-helical conformation at low temperatures and a random-coil conformation at high temperatures. Increasing the hydrophobic interaction strength to $R = 1/8, 1/6$ and $1/4$ induces a second transition to a β-structure, resulting in an α-helical conformation at low temperatures, a β-structure at intermediate temperatures, and a random-coil at high temperatures. At very high values of the hydrophobic interaction strength, i.e. $R \geq 1/2$, polyalanines become β-structures at low temperatures and random-coils at high temperatures. At low and intermediate values of hydrophobic interaction strength, i.e. $R < 1/2$, the transition temperatures between an α-helical
state to a non-α-helical state are between $T^* = 0.085$ and $T^* = 0.11$. These results are qualitatively in good agreement with experiments which show that polyalanine adopts an α-helical conformation in hydrophobic environments such as the solid state or in non-polar organic solutions and a β-structure conformation in polar aqueous solution $^{99-103, 86, 104}$. This is similarly observed in experiments on many heterogeneous peptides which can be folded into alternative stable structures by changing the solution conditions such as the pH, salt or organic cosolvent concentration, peptide concentration, and the redox state $^{105-119}$. In this paper, although simulations were conducted at $R = 1/10$, 1/8, and 1/6, we focus most of our analysis on the $R = 1/10$ simulations.

### 4.3.1 Fibril formation as a function of temperature and concentration

In this section, we present the results from our fibril formation simulations of 48-peptide systems initially in a random-coil configuration. Simulations are conducted at concentrations $c = 0.5, 1, 2.5, 5, 10,$ and $20$ mM, which range from the very dilute regime, in which most peptides do not interact with neighboring peptides, to the highly concentrated regime, in which most peptides are in contact with neighboring peptides. At each concentration, simulations are performed at temperatures $T^* = 0.08$, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14, and 0.15, which ranges from temperatures that are well below the folding temperature for a single α-helix ($T^* = 0.11$) to a temperature ($T^* = 0.15$) that is so high that the peptides cannot form or maintain hydrogen bonds or hydrophobic interactions, and are in random-coil conformations. The parameter $R = \epsilon_{HP}/\epsilon_{HB}$ is set at 1/10. The results at each concentration and temperature are taken from the end of at least three different runs. If the potential energy of the system from three initial runs varied by more than 2.5%, then up to three more runs were conducted as warranted.
The percentage of peptides in each structure of interest is highly dependent on environmental conditions such as the temperature and the peptide concentration. In terms of forming α-helices, it is evident from Figure 4.2, which plots the percentage of peptides that form α-helices versus temperature at different concentrations, that there is an inverse relationship between the percentage of peptides that form this structure and the concentration. Alpha-helices are unlikely to be formed at concentrations greater than or equal to 5mM. There is an optimal range of temperatures for forming α-helices; at concentrations that are less than 5mM, the temperature at which formation of α-helices is at a maximum is \( T^* = 0.09 \). The maximum percentage of peptides observed to form α-helices in this study is about 65% and occurs at \( T^* = 0.09 \) and \( c = 0.5 \text{mM} \). At low temperatures (\( T^* < 0.09 \)) and low concentrations (\( c < 5 \text{mM} \)), the system is kinetically trapped; in this case, the system forms either isolated \( \beta \)-structures (data not shown) or aggregates that are mostly amorphous (discussed below) instead of α-helices. At temperatures higher than 0.09 and low concentrations (\( c < 5 \text{mM} \)), the percentage of peptides that form α-helices decreases as the temperature increases, ending at a minimum of zero percent at \( T^* = 0.12 \). As the concentration decreases from \( c = 2.5 \text{mM} \) to 1.0mM and then 0.5mM, the midpoint between the maximum and the minimum percentages of peptides that form α-helices at each concentration gets close to \( T^* = 0.11 \), the midpoint of the folding transition (50% helicity) of a single peptide from our previous simulations \(^7\). As the system gets diluted beyond \( c = 0.5 \text{mM} \), these two transition midpoints are expected to converge.

Aggregation is relatively high for all concentrations except \( c = 0.5 \text{mM} \) over a relatively wide range of temperatures spanning from \( T^* = 0.08 \) to 0.14 as shown by Figure 4.3, which plots the percentage of peptides in (all types of) aggregates as a function of temperature at various con-
centrations. At high concentrations, $c=5-20\text{mM}$, 100% of the peptides in the system form aggregates. As the concentration decreases to $c=2.5\text{mM}$, the aggregation percentage decreases slightly at intermediate temperatures ($T^*=0.10-0.11$) and somewhat more significantly at low temperatures ($T^*=0.08-0.09$). As the concentration decreases further to $c=1\text{mM}$, the aggregation percentage decreases to 50% at low temperatures ($T^*=0.08$) and 75% at high temperatures ($T^*=0.13$). As the concentration decreases to $c=0.5\text{mM}$, the aggregation percentage decreases significantly to 10% at low temperatures ($T^*=0.08$) and 30% at high temperatures ($T^*=0.12$).

There is a conformational transition from $\alpha$-helices at low temperatures to $\beta$-sheets at high temperatures for systems at intermediate concentrations ($c=1-2.5\text{mM}$) as can be seen by comparing the results in Figure 4.2 and Figure 4.4, which depicts the percentage of peptides in $\beta$-sheets as a function of temperature at different concentrations. For example, at $c=1\text{mM}$ and $T^*=0.09$, 55% of the peptides are $\alpha$-helical and 5% of the peptides are in $\beta$-sheets but at $T^*=0.13$, none of the peptides are $\alpha$-helical and 75% of the peptides are in $\beta$-sheets. The transition from $\alpha$-helical structures to $\beta$-sheet complexes, which is similarly observed in vitro by Blondelle et al. for polyalanine peptides $^{85,86}$, mimics the structural transition believed to be a prerequisite for fibril nucleation and growth in general $^{87-89,90,47,91}$. At low concentrations ($c=0.5\text{mM}$), the system forms relatively few $\beta$-sheet peptides; in fact, the maximum percentage of peptides that form $\beta$-sheets at these low concentrations is only about 20% at high temperatures ($T^*=0.12$). At high concentrations ($c=5-20\text{mM}$), the system forms a very high number of $\beta$-sheet peptides over a relatively large range of temperatures. When comparing the results in Figure 4.3 and 4.4 for systems at $c=5-10\text{mM}$, it is evident that most aggregated peptides are in $\beta$-sheets.

Fibril formation is seen only at high temperatures and high concentrations as indicated in
Figure 4.5, which depicts the percentage of peptides in fibrils as a function of temperature at different concentrations. At high concentrations, \( c = 5, 10 \) and 20mM, the dependency of fibril formation on the temperature is similar; fibril formation increases as the temperature increases up to \( T^* = 0.13 \) then decreases as the temperature increases further. However, as the concentration is increased from \( c = 5 \) mM to \( c = 20 \) mM, the maximum in the percentage of peptides in fibrils broadens to include lower temperatures. For example, the temperature range for the \( c = 5 \) mM system within which fibril formation is high is between \( T^* = 0.13 \) and \( T^* = 0.12 \). The lower limit of this range moves to \( T^* = 0.11 \) for the \( c = 10 \) mM system, and then to \( T^* = 0.10 \) for the \( c = 20 \) mM system. In other words, the critical temperature for forming fibrils decreases with peptide concentration, which is similarly observed in vitro by Blondelle et al. for polyalanine peptides. The maximum percentage of peptides observed to form fibrils in this study is about 90% and occurs at \( T^* = 0.10 - 0.13 \) for the \( c = 20 \) mM system. A comparison of Figures 4.4 and 4.5 shows that only a portion of the \( \beta \)-sheet peptides form fibrils. At low temperatures \( (T^* = 0.08 - 0.09) \) and high concentrations \( (c = 5 - 10 \text{mM}) \), the \( \beta \)-sheets tend to stick together at a relatively large angle (i.e greater than 35°) resulting in a low yield of fibrils. At high temperatures \( (T^* = 0.13 - 0.14) \) and intermediate concentrations \( (c = 1 - 2.5 \text{mM}) \), the \( \beta \)-sheets tend to be isolated without forming fibrils.

The formation of amorphous aggregates is seen mostly at intermediate concentrations \( (c = 1 - 2.5 \text{mM}) \) and low temperatures \( (T^* = 0.08 - 0.09) \) as shown in Figure 4.6, which plots the percentage of peptides in amorphous aggregates as a function of temperature at various concentrations. This figure indicates that the highest percentage of the peptides that form amorphous aggregates is around 50% at \( c = 2.5 \) mM. The aggregates formed at this condition contain \( \alpha \)-helices as can be seen by comparing the results in Figures 4.2 and 4.3 for systems at \( c = 1.0 \), and 2.5mM and at...
$T^*=0.08-0.09$. For example, at $c=2.5mM$ about 30% of the peptides in the system form either isolated or attached $\alpha$-helices while 90% of the peptides form aggregates; at least 20% of the peptides that are $\alpha$-helices reside inside aggregates (data not shown). The $\alpha$-helices that reside inside of aggregates are more stable than the isolated $\alpha$-helices since each aggregated $\alpha$-helix can form multiple extra hydrophobic interactions with neighboring peptides. A picture of one such amorphous aggregate is shown in Figure 4.7 in which $\alpha$-helices are shown in blue.

The systems that form amorphous aggregates at intermediate concentrations and low temperatures are believed to be kinetically trapped in local minima. At low temperatures, peptides tend to form isolated $\alpha$-helices at low concentrations and $\beta$-sheets at high concentrations. At intermediate concentrations, peptides tend to first form isolated $\alpha$-helices which then cluster to form aggregates. These $\alpha$-helices try to unfold to form ordered aggregates; however, at low temperatures, they become trapped in amorphous aggregates.

### 4.3.2 Key fibril formation events

Experimental studies of fibril formation strongly suggest that it is a nucleation-dependent process $^{120-123}$, with small micellar $^{124}$ or amorphous $^{125,122,123}$ aggregates preceding nucleus formation. Based on a recent series of experiments on a yeast prion protein Sup35, Serio et al. describe fibril formation as a nucleated conformational conversion process in which the formation of small amorphous aggregates containing only 20 to 80 protein monomers precedes critical nucleus formation and subsequent rapid growth of large fibrils $^{122}$. While our simulation system is quite small, we observe the same trend of going from amorphous aggregates to ordered nucleus and finally to stable fibril as described below, first by presenting snapshots of our simulations.
and then by analyzing our simulation data.

Figure 4.8 shows snapshots of the fibrillization process, which were taken at various reduced times, $t^*$, for the 48-peptide simulation at $R=1/10$, $T^*=0.10$, and a peptide concentration of 20mM. All hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets once the fibril is formed. Starting in random coil conformations at reduced time $t^*=0$, the peptides begin to form small aggregates almost immediately. By $t^*=6.4$, one aggregate becomes a two-peptide $\beta$-sheet (i.e., the white sheet at the upper left edge of the box which wraps to the lower left side of the box), one aggregate becomes a three-peptide $\beta$-sheet (i.e., the purple sheet at the lower left corner of the box), and another aggregate becomes a four-peptide $\beta$-sheet (i.e., the light blue sheet at the lower right corner of the box). At $t^*=11.5$, most peptides are in aggregates that are somewhat $\beta$-sheet-like. The big aggregate contains some internal fibrillar structure: a two-peptide $\beta$-sheet (white), a three-peptide $\beta$-sheet (purple) and a four-peptide $\beta$-sheet (light blue). By $t^*=21.8$, this fibrillar structure has grown into a three-sheet fibril; in addition, there is an isolated four-peptide sheet (dark blue) that is about to latch onto the fibrillar structure. At $t^*=24.7$, the isolated four-peptide sheet (dark blue) in the previous slide has become part of the fibril which now contains four $\beta$-sheets and some of the monomeric peptides have associated to form a separate $\beta$-sheet (green). By $t^*=51.8$, both the fibrillar structure and the unattached $\beta$-sheet (green) have grown by adding peptides to the ends of $\beta$-sheets. By $t^*=68.1$, the unattached $\beta$-sheet (green) has latched onto the four-sheet fibril. By $t^*=89.8$, the fibrillar structure is stable and consists of five $\beta$-sheets, each with seven or eight peptides. Even after a long equilibration time, $t^*=140.1$, the fibrillar structure remains stable. This fibrillization process exhibits a sequence of events that is typical of
many of our simulations at different peptide concentrations and temperatures. In essence, at the
beginning of the fibrillization process, the system of denatured peptides stays in a lag phase
during which some amorphous aggregates form. These aggregates then convert themselves into
small $\beta$-sheets containing aligned $\beta$-strands. Once these $\beta$-sheet structures attain a certain size,
they come together and align one-by-one, creating a fibril. This raises the question of whether
the nucleus is a stretched-out peptide (a $\beta$-strand), a dimer, a trimer or a much larger structure.

Quantitative analysis of our simulation data confirms that our simulations exhibit the nu-
cleated conformational conversion from amorphous aggregates to small fibrillar structures with
subsequent rapid growth of large fibrils suggested by Serio et al.\textsuperscript{122}. This can be seen in Figure
4.9, which plots the percentage of peptides in aggregates, $\beta$-sheets, and fibrillar structures and
the average number of peptides per $\beta$-sheet per fibril and the average number of $\beta$-sheets per
fibril over time $t^*$. The data here is taken from the simulation shown in Figure 4.8. As indicated
in Figure 4.9(a), aggregates are formed instantaneously. There is a relatively small delay time
of 3 reduced time units before some of the aggregates convert into $\beta$-sheets and a lag time of 13
reduced time units before fibrils start to appear. As indicated in Figure 4.9(b), the early fibrils
are relatively small, containing two $\beta$-sheets each consisting of three or four peptides. The num-
ber of $\beta$-sheets per fibril increases over time indicating that the fibril grows in part by adding
already-formed $\beta$-sheets to its side. In addition, the number of peptides per $\beta$-sheet increases
gradually with time, indicating that the fibril also grows by adding peptides to the end of each
$\beta$-sheet, thereby lengthening along the fibril axis. Once the fibril reaches its final size of five
$\beta$-sheets (the number of $\beta$-sheets per fibril remains constant), the number of peptides per $\beta$-sheet
per fibril continues to increase from 6 to 8 peptides. In summary, there are two growth mecha-
nisms once the nucleus has been formed: the first is by adding already-formed $\beta$-sheets to the side and the second is by adding individual peptides to the end of the fibril. The trends displayed in Figure 4.9 are typically seen in all our simulations at different temperatures and concentrations. The kinetics of fibril formation will be explored in more detail and the critical nucleus will be identified in a future publication.

The forces that are responsible for the early events in bringing peptides together to form amorphous aggregates are the hydrophobic interactions as indicated by Figure 4.10, which plots the average number of hydrogen bonds and hydrophobic interactions as a function of time, $t^*$, from multiple simulations on the 48-peptide system at $c=20\text{mM}$ and $T^*=0.10$. This figure shows that there is a disproportionately high number of hydrophobic interactions compared to the number of hydrogen bonds that form at the beginning of simulations. These hydrophobic interactions are responsible for the formation of amorphous aggregates which then convert themselves into $\beta$-sheets and fibrils as described above. The role of the hydrophobic interactions in the assembly of peptides into ordered structures is also observed in a simulation study by Klimov et al.\textsuperscript{44} who found that the substitution of hydrophobic residues with polar ones destabilizes ordered $\alpha\beta_{16-22}$ oligomers and decreases $\beta$ strand propensities.

### 4.3.3 Fibril stability

In our fibril stability simulations, fibrils are stable at temperatures above those at which the $\alpha$-helical model peptides are stable as indicated in Figure 4.11, which is a plot of the fibril order parameters, $Q_{\text{intra-sheet}}$ and $Q_{\text{inter-sheet}}$, as a function of time during the high temperature stability simulations at $T^*=0.14$, 0.15, 0.16, 0.17 and 0.18. The quantity $Q_{\text{intra-sheet}}$ is defined
as the sum of the numbers of intra-sheet hydrogen bonds and intra-sheet hydrophobic interactions within those \( \beta \)-sheets present in the original fibril that are present at time \( t^* \) divided by the sum of their initial numbers. In other words, \( Q_{\text{intra-sheet}} \) measures the number of hydrogen bonds and hydrophobic interactions broken and formed on the peptides within each original \( \beta \)-sheet and hence represents the intra-sheet stability. The quantity \( Q_{\text{inter-sheet}} \) is defined as the number of inter-sheet hydrophobic interactions between \( \beta \)-sheets present in the original fibril that are present at time \( t^* \) divided by their initial number; \( Q_{\text{inter-sheet}} \) measures the number of inter-sheet hydrophobic interactions broken and formed and hence represents the inter-sheet stability.

In each simulation, the starting configuration is the structure formed at the end of a simulation at \( T^*=0.11 \) which itself started out as a system of random coils. The system is instantaneously heated to the temperature of interest and thereafter the temperature remains constant. The order parameters for the initial structures are unity (\( Q_{\text{intra-sheet}} \) and \( Q_{\text{inter-sheet}} = 1 \)). At the lowest temperatures in this range \( T^*=0.12-0.13 \), the fibril system stays in the original structure for extended times (data not shown). As seen in Figure 4.11, the fibril continues to maintain its stability over extended times at \( T^*=0.14 \).

The fibril can recover its original stability along the fibril axis more easily than perpendicular to the fibril axis when exposed to high temperatures as indicated by the values of the order parameters \( Q_{\text{intra-sheet}} \) and \( Q_{\text{inter-sheet}} \) at \( T^*=0.15 \). In this case, the fibril gradually loses a small amount of its original intra-sheet stability over long periods of time; on the other hand, the fibril loses a small amount of its original inter-sheet stability only at the beginning of the simulations but then regains and even strengthens its inter-sheet stability toward the end of the simulations. The fibril gains this additional inter-sheet stability by shifting each \( \beta \)-sheet a small distance so as
to maximize the number of inter-sheet hydrophobic contacts. At a slightly higher temperature, $T^*=0.16$, the fibril loses approximately 45% of its original intra-sheet contacts at long times. Although the fibril initially loses approximately 30% of its original inter-sheet contacts, it regains some of its inter-sheet stability; after long times, it has lost only 20% of its original inter-sheet contacts. At $T^*=0.17-0.18$, the fibril quickly loses most of its original contacts and disperses into random coils.

Our fibrillar complexes are stable at higher temperatures than the isolated $\alpha$-helices are stable, which is similar to experimental results on polyalanine-based peptides\textsuperscript{85,103}. The temperature at which the fibril starts to lose a significant fraction of its original contacts in our simulations as indicated in Figure 4.11 is $T^*=0.16$, which is well above the unfolding temperature for isolated $\alpha$-helices ($T^*=0.11$) from a previous study\textsuperscript{79}. In our systems, the model fibrils are more stable than the model isolated $\alpha$-helices at high temperatures because model fibrils contain more hydrogen bonds. Each peptide within a $\beta$-sheet of a fibril can form up to 15 inter-peptide hydrogen bonds to each of its two neighboring peptides and zero intra-peptide hydrogen bonds. This can be compared to just 12 intra-peptide hydrogen bonds and zero inter-peptide hydrogen bonds formed by an isolated $\alpha$-helix. Moreover, inter-peptide hydrophobic contacts, which are numerous in systems of model fibrils, are not present in diluted systems of model $\alpha$-helices which rarely interact with each other.

Fibril growth takes place at higher temperatures than fibril formation as indicated in Figure 4.12, which plots the average number of peptides per $\beta$-sheet per fibril and the average number of $\beta$-sheets per fibril over time $t^*$ from fibril stability simulations at $T^*=0.15$ and $T^*=0.16$. Compared with Figure 4.5, which shows that fibril formation happens only at temperatures that
are up to $T^*=0.14$, the already-formed fibril in Figure 4.12 continues to grow at $T^*=0.15$ by adding monomeric peptides that are still available in the system to the ends of the $\beta$-sheets. This suggests that fibril formation requires the presence of a nucleus especially at high temperatures. Such nucleation will be explored in more detail in a future publication.

The fibril is more stable along the fibril axis than perpendicular to it at high temperatures as indicated in Figure 4.12 for $T^*=0.16$. When instantaneously exposed to high temperature, the fibril quickly loses more than 30% of its original number of $\beta$-sheets, which then disintegrate into monomeric peptides (data not shown). On the other hand, peptides in the remaining $\beta$-sheets not only maintain their numbers, they continue to grow slightly.

### 4.3.4 Fibril structure

The fibrils observed in our simulations mimic the structural characteristics of fibrillar aggregates currently described in the literature. Specifically, most peptides in our fibrils are arranged in a parallel orientation. In fact, the percentage of peptides whose C-terminus are on a particular side of their $\beta$-sheets in the fibril (with N-terminus on the other side) compared with the total number of peptides in all $\beta$-sheets in the fibril is 93.3% (±5.7) for systems at high temperatures ($T^*=0.12-0.14$). In addition, alanine sidechains of adjacent peptides within each $\beta$-sheet are aligned directly as can be seen in Figure 4.13, which displays a close-up snapshot of the 48-peptide fibrillar structure formed at $R=1/10$, $c=5.0\,\text{mM}$, and $T^*=0.13$. Also, the end residues of neighboring peptides are relatively aligned directly, i.e. in register. In-register parallel $\beta$-sheets are similarly observed in experiments on fibrils formed by $\beta$-amyloid proteins \cite{126-128} and in a computer simulation study by Gsponer et al.\cite{39} on a three-peptide $\beta$-sheet formed by the
heptapeptide GNNQQNY from the yeast prion Sup35. Moreover, Kuwata et al.\textsuperscript{38} performed both NMR-detected hydrogen bond exchange experiments and molecular dynamics simulations on a fibril-forming mouse prion protein fragment 106-126, and observed that parallel $\beta$-sheets, which were stabilized by the central alanine-based residues VAGAAAAGAV, are the most stable structures. The predominance of parallel pairing in our model differs from the results of a previous study by Ma et al.\textsuperscript{32} who examined the stability of proposed fibril structures by conducting nanosecond simulations at high temperatures on already-formed parallel and antiparallel fibril oligomers containing either AGAAAAGA or AAAAAAAA peptides and found that the antiparallel orientation is the most stable structure. Each of the peptides in our fibrils is almost fully extended as indicated by its end-to-end distance and radius of gyration, which are 40.81Å ($\pm$4.76) and 13.10Å ($\pm$1.04) compared with 51.24Å and 15.10Å, respectively, for the completely stretched-out linear peptide conformation and 29.05Å ($\pm$0.39) and 10.45 Å ($\pm$0.08), respectively, for the random-coil conformation.

The $\beta$-sheets in the fibrillar structures that result from our simulations are stabilized by intra-sheet backbone-backbone hydrogen bonds running parallel to the length of the sheet. The adjoining $\beta$-sheets are held together by inter-sheet sidechain-sidechain hydrophobic interactions oriented nearly perpendicular to the length of the sheet. In addition, each $\beta$-sheet is stabilized by intra-sheet hydrophobic interactions oriented parallel to the length of the sheet. These intra-sheet hydrophobic contacts are a consequence of the tight packing in the complex and serve to lower the overall internal energy of the complex. The packing of our fibrillar structures mimics the close packed sheets found in experiments on a variety of peptides \textsuperscript{129,121,11}. Our typical fibrillar complex is a dense, ordered structure with tightly packed sidechains. The intra-sheet ($C_\alpha$ to $C_\alpha$)
distance is 5.05 Å (±0.07), which is slightly higher than the experimentally determined value of 4.7-4.8 Å for close packed sheets of different peptides.\textsuperscript{129,121,11} In addition, the inter-sheet ($C_{\alpha}$ to $C_{\alpha}$) distance is 7.5 Å (±0.5), which is comparable to the low end of experimental values of 8-10 Å for transthyretin fragments\textsuperscript{130} and significantly smaller than the experimentally determined value of ≈ 9-10 Å for close packed sheets of Alzheimer’s peptides.\textsuperscript{129,121,11} This is not surprising since both transthyretin fragments and β-amyloid peptides have big bulky sidechains whereas our model peptides consist mostly of alanine residues which have one of the smallest size sidechains. As a result, the alanine sidechains on one β-sheet lie between the neighboring alanine sidechains on the adjacent β-sheet as can be seen in Figure 4.13.

Our fibrils contain the same number of β-sheets as in real fibrils, four to six sheets.\textsuperscript{11,96–98} Snapshots of fibrils formed for different system sizes are shown in Figure 4.14. At $c=5\text{mM}$, $R=1/10$, and $T^*=0.13$, the fibrils formed from the 12-peptide system contain between two to three β-sheets whereas the fibrils from the 24-peptide system contain between two to four β-sheets. At the same concentration, temperature, and hydrophobic interaction strength, the 48-peptide system forms fibrils containing between three to six β-sheets. When the system size is increased further to 96 peptides, the fibrils again contain between four to six β-sheets. This indicates that as the system size is increased significantly, the number of β-sheets asymptotes to a value near six. Once the fibrillar structure reach its critical β-sheet number, monomeric peptides tend to attach to the fibrillar structure rather than creating an isolated β-sheet. The kinetics of bigger systems will be explored in a future publication.
4.3.5 **Fibril formation as a function of the hydrophobic interaction strength between non-polar sidechains**

The strength of the hydrophobic interaction between non-polar sidechains plays an important role in fibril formation as indicated in Figure 4.15, which displays the percentage of peptides in (a) fibrillar and (b) $\beta$-sheet structures as a function of temperature for the c=5mM system at different ratios, $R \equiv \epsilon_{HP}/\epsilon_{HB}$, of the strength of the hydrophobic interaction relative to that of the hydrogen bond. As the strength of the hydrophobic interaction increases from $R=1/10$ to $R=1/8$, the percentage of peptides in fibrils slightly decreases at all temperatures as seen in Figure 4.15(a). But as the strength of the hydrophobic interaction increases from $R=1/8$ to $R=1/6$, the percentage of peptides in fibrils decreases dramatically at all temperatures. Instead, the system at $R=1/6$ forms amorphous aggregates. The average number of inter-peptide hydrophobic interactions in amorphous aggregates is 52% greater than in fibrils while the number of inter-peptide hydrogen bonds in amorphous aggregates is 35% less than in fibrils (data not shown). This indicates that having a large hydrophobic interaction strength essentially overwhelms the system, forcing it to form more hydrophobic interactions than hydrogen bonds and hence favoring the formation of amorphous aggregates.

Fibril formation seems to be more sensitive than $\beta$-sheet formation to the strength of the hydrophobic interaction between non-polar sidechains as can be seen by comparing the results in Figures 4.15(a) and 4.15(b). Although there are no fibrils formed at $R=1/6$, about half of the peptides form $\beta$-sheets which are intertwined in large aggregates. However, unlike the fibrils, whose average $\beta$-sheet size is about 7 peptides (data not shown), these $\beta$-sheets are relatively small, each containing approximately two peptides (data not shown). These $\beta$-sheets are twisted
around each other to maximize the number of hydrophobic interactions, as discussed above. A picture of one such aggregate is shown in Figure 4.16.

4.4 Conclusions

Computer simulations offer unique opportunities to observe and analyze molecular-level events in protein systems that are difficult or impossible to observe experimentally. In simulating large multi-peptide systems using an intermediate resolution protein model in conjunction with the discontinuous molecular dynamics, we have been able to observe the formation of fibrils containing polyalanine peptides starting from random coils. We find that there is a strong relationship between the formation of $\alpha$-helices, $\beta$-sheets, aggregates and fibrils and the environmental conditions such as temperature, concentration, and hydrophobic interaction strength. At low peptide concentrations, random-coil peptides form $\alpha$-helices at low temperatures and revert to random-coil conformations at high temperatures. When the concentration is increased slightly to intermediate values, random-coil peptides assemble into $\alpha$-helices at low temperatures and large $\beta$-sheet structures at high temperatures. As the concentration is increased further to relatively high values, random-coil peptides forms $\beta$-sheets over a wide range of temperatures; these $\beta$-sheets assemble into fibrils at temperatures above a critical temperature that decreases with concentration. Fibrils are stable at temperatures above those at which the $\alpha$-helical model peptides are stable. At very high temperatures and all concentrations, the system is in a random-coil state. These results agree qualitatively with the experimental results of Blondelle et al. on Ac-KA$_{14}$K-NH$_2$ peptides $^{85,86}$. They observed monomeric $\alpha$-helical structures at 100$\mu$M and 25°C. As the peptide concentration increased to 1mM, they found that $\beta$-sheet complex formation in-
creased with increasing temperature, exhibiting an S-shaped dependence of temperature with a
critical temperature of 65°C. As the peptide concentration increased to 1.8mM, they found that
the critical temperature at which β-sheets start to form decreased to 45°C.

We observe that fibril formation proceeds from amorphous aggregates to ordered nucleus
and finally to stable fibrils. These findings are similar to those in experimental studies which
strongly suggest that fibril formation is a nucleation-dependent process involving a con-
version from amorphous aggregates to small fibrillar structures. Once the nucleus has
formed, our fibril growth involves both elongation in which monomeric peptides are added to
the end of β-sheets and lateral growth in which already-formed β-sheets are added to the side of
the fibril. The fibrils observed in our simulations mimic the structural characteristics observed
in experiments in that most peptides in our fibrils are arranged in a parallel orientation,
the intra-sheet (Cα to Cα) distance is 5.05Å (±0.07) (compared to experimental values of 4-7-
4.8Å), the inter-sheet (Cα to Cα) distance is 7.5Å (±0.5) (compared to experimental
values of 8-10Å), and our fibrils contain about six β-sheets with each containing
multiple peptides. Finally, we find that when the strength of the hydrophobic interaction
between non-polar sidechains is high compared to the strength of the hydrogen bond interaction,
amorphous rather than fibrillar aggregates are formed.

Based on this initial success in simulating the formation of fibrils, we expect that computer
simulations can assist in the identification of sources of aggregate stability, information vital to
the rational design of therapeutic strategies in protein aggregation diseases. It is also likely that
computer simulation studies will lead to a better understanding of the fundamental mechanisms
of aggregate formation, permitting a clearer picture of aggregate involvement in the pathology
of amyloid diseases.

It is important to point out that our model and analysis are subject to a number of limitations. First, we do not include charged residues at the ends of the model peptide chains, which have been shown to be important in experimental systems for reducing amorphous aggregation and precipitation. Second, it is possible that a more elaborate model force field is required to adequately represent peptides and their environment. Specifically, we have neglected solvent effects by incorporating solvent as a potential of mean force which has an equal effect on all chain united atoms regardless of the chain conformation. Consequently united atoms in the interior of a compact structure are just as affected by solvent as if they were fully exposed in a random coil structure. A more accurate solvation model would allow forces, such as the hydrogen bonding force in the core of a collapsed chain or an aggregate to be different from those at the surface. In addition, we have fixed the strengths of the hydrogen bond and hydrophobic interactions relative to temperature. Dill et al.\textsuperscript{131} and Chan et al.\textsuperscript{132} have proposed a temperature-dependent hydrophobic potential that undergoes a maximum at intermediate temperatures, accounting for weakened interactions from cold denaturation at low temperature and from heat denaturation at high temperature \textsuperscript{131}. Further simulation studies with our model will be required to probe the importance of temperature-dependent interactions. Despite these limitations, our results agree very well with experiments on fibril formation.

4.5 Acknowledgments

The authors are grateful to Sylvie Blondelle, Ken Dill, Angel Garcia, Jeffery Kelly, and Dan Kirschner for helpful discussions. This work was supported by the National Institutes of
Health under grant number GM-56766 and the National Science Foundation under grant number CTS-9704044.
4.6 References


[103] D. K. Lee and A. Ramamoorthy, “Determination of the solid-state conformations of


4.7 Figures

4.1 Covalent bonds are shown with narrow black lines connecting united atoms. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive Cα distances, and residue L-isomerization. Note that the united atoms are not shown full size for ease of viewing. ................................................. 173

4.2 The percentage of peptides in the α-helical conformation obtained at the end of simulations at different temperatures and concentrations. ..................... 174

4.3 The percentage of peptides in aggregates obtained at the end of simulations at different temperatures and concentrations. ......................... 175

4.4 The percentage of peptides in β-sheet structures obtained at the end of simulations at different temperatures and concentrations. ....................... 176

4.5 The percentage of peptides in fibrillar structures obtained at the end of simulations at different temperatures and concentrations. ....................... 177

4.6 The percentage of peptides in amorphous structures obtained at the end of simulations at different temperatures and concentrations. ....................... 178

4.7 A snapshot of an aggregate obtained from a simulation at \( R=1/10, c=2.5 \text{mM}, \) and \( T^* = 0.08 \). Alpha-helices are shown in dark blue. ......................... 179
4.8 Snapshots of the 48-peptide system at various times during a simulation that results in the formation of a fibril. All hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets once the fibril is formed.

4.9 (a) The percentage of peptides in aggregates (of any type), $\beta$-sheets, and fibrillar structures and (b) the average number of peptides per $\beta$-sheet in a fibril and the average number of $\beta$-sheets per fibril versus reduced time, $t^*$, for the 48-peptide system at $c=20\text{mM}$ and $T^*=0.10$.

4.10 The average number of hydrogen bonds and hydrophobic interactions versus reduced time, $t^*$, for the 48-peptide system at $c=20\text{mM}$ and $T^*=0.10$.

4.11 Fibril stability parameters $Q_{\text{intra-sheet}}$ and $Q_{\text{inter-sheet}}$ versus reduced time, $t^*$ from simulations starting with fibrils formed at $T^*=0.11$ performed at $T^*=0.14$ (filled circles), $T^*=0.15$ (open squares), $T^*=0.16$ (filled diamonds), $T^*=0.17$ (open triangles), and $T^*=0.18$ (filled triangles).

4.12 The average number of peptides per $\beta$-sheet in a fibril (filled circles) and the average number of $\beta$-sheets per fibril (open squares) versus reduced time, $t^*$, for the 48-peptide system at $T^*=0.15$ (top panel) and $T^*=0.16$ (bottom panel). Results are from simulations starting with fibrils formed at $T^*=0.11$ that are subsequently heated to the temperatures of interest.
4.13 A close-up snapshot of the 48-peptide fibrillar structure viewed down the fibril axis, formed at $R=1/10$, $c=5.0\text{mM}$, and $T^*=0.13$. Note that all hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets. United atoms are not shown full size for ease of viewing.

4.14 Snapshots of fibrillar structures formed at different system sizes at $R=1/10$, $c=5.0\text{mM}$, and $T^*=0.13$.

4.15 The percentage of peptides in (a) fibrillar structures and (b) $\beta$-sheet structures at concentration $c=5\text{mM}$ versus temperature at three ratios $R$ of the strength of the hydrophobic interaction relative to that the hydrogen bond: $R=1/10$ (filled circles), $R=1/8$ (open squares), and $R=1/6$ (filled diamonds).

4.16 A snapshot of an aggregate obtained from a simulation at $R=1/6$, $c=5.0\text{mM}$, and $T^*=0.13$. 
Figure 4.1: Covalent bonds are shown with narrow black lines connecting united atoms. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_\alpha$ distances, and residue L-isomerization. Note that the united atoms are not shown full size for ease of viewing.
Figure 4.2: The percentage of peptides in the $\alpha$-helical conformation obtained at the end of simulations at different temperatures and concentrations.
Figure 4.3: The percentage of peptides in aggregates obtained at the end of simulations at different temperatures and concentrations.
Figure 4.4: The percentage of peptides in β-sheet structures obtained at the end of simulations at different temperatures and concentrations.
Figure 4.5: The percentage of peptides in fibrillar structures obtained at the end of simulations at different temperatures and concentrations.
Figure 4.6: The percentage of peptides in amorphous structures obtained at the end of simulations at different temperatures and concentrations.
Figure 4.7: A snapshot of an aggregate obtained from a simulation at $R=1/10$, $c=2.5$ mM, and $T^*=0.08$. Alpha-helices are shown in dark blue.
Figure 4.8: Snapshots of the 48-peptide system at various times during a simulation that results in the formation of a fibril. All hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets once the fibril is formed.
Figure 4.9: (a) The percentage of peptides in aggregates (of any type), $\beta$-sheets, and fibrillar structures and (b) the average number of peptides per $\beta$-sheet in a fibril and the average number of $\beta$-sheets per fibril versus reduced time, $t^*$, for the 48-peptide system at $c=20\text{mM}$ and $T^*=0.10$. 
Figure 4.10: The average number of hydrogen bonds and hydrophobic interactions versus reduced time, $t^*$, for the 48-peptide system at $c=20\text{mM}$ and $T^\ast=0.10$. 
Figure 4.11: Fibril stability parameters $Q_{\text{intra-sheet}}$ and $Q_{\text{inter-sheet}}$ versus reduced time, $t^*$ from simulations starting with fibrils formed at $T^*=0.11$ performed at $T^*=0.14$ (filled circles), $T^*=0.15$ (open squares), $T^*=0.16$ (filled diamonds), $T^*=0.17$ (open triangles), and $T^*=0.18$ (filled triangles).
Figure 4.12: The average number of peptides per β-sheet in a fibril (filled circles) and the average number of β-sheets per fibril (open squares) versus reduced time, $t^*$, for the 48-peptide system at $T^*=0.15$ (top panel) and $T^*=0.16$ (bottom panel). Results are from simulations starting with fibrils formed at $T^*=0.11$ that are subsequently heated to the temperatures of interest.
Figure 4.13: A close-up snapshot of the 48-peptide fibrillar structure viewed down the fibril axis, formed at $R = 1/10$, $c = 5.0\text{mM}$, and $T^* = 0.13$. Note that all hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets. United atoms are not shown full size for ease of viewing.
Figure 4.14: Snapshots of fibrillar structures formed at different system sizes at $R=1/10$, $c=5.0\text{mM}$, and $T^*=0.13$. 

(a) 12 peptides  
(b) 24 peptides  
(c) 48 peptides  
(d) 96 peptides
Figure 4.15: The percentage of peptides in (a) fibrillar structures and (b) β-sheet structures at concentration $c=5$ mM versus temperature at three ratios $R$ of the strength of the hydrophobic interaction relative to that the hydrogen bond: $R=1/10$ (filled circles), $R=1/8$ (open squares), and $R=1/6$ (filled diamonds).
Figure 4.16: A snapshot of an aggregate obtained from a simulation at $R=1/6$, $c=5.0\text{mM}$, and $T^*=0.13$. 
5.1 Introduction

Ordered β-sheet complexes are the underlying structural components of the intra- and extracellular fibrillar protein deposits, termed amyloid fibrils, that are associated with a variety of human diseases, including Alzheimer’s, Parkinson’s and the prion diseases \(^1\text{–}^6\). Though extensively studied, the mechanisms that govern the formation of amyloid fibrils have not yet been fully determined. However, recent evidence that proteins other than those associated with amyloid diseases form fibrils \textit{in vitro} under mildly denaturing conditions \(^7\text{–}^9\) has led leaders in the field to suggest that fibril formation is an intrinsic property of polypeptides, albeit under appropriate conditions. This implies that progress towards understanding the origins of various protein deposition diseases can be made by \textit{in vitro} or \textit{in silico} examination of the general features of protein fibrillization using model proteins that are less complex than the specific amyloidogenic protein. Here we perform computer simulations on systems containing polyalanine-based peptides of the sequence Ac-KA\textsubscript{14}K-NH\textsubscript{2} in an attempt to help further understanding of the molecular-
level mechanisms that are responsible for fibril formation. Our molecular dynamics simulations are conducted on systems containing many peptides initially in the random-coil state.

Four mechanisms have been proposed to describe the conformational transformation and assembly of normally soluble proteins into ordered aggregates \(^\text{10}\). The first is the templated assembly (TA) mechanism \(^\text{11,12}\), in which a soluble random-coil peptide binds to a pre-assembled \(\beta\)-sheet-rich nucleus and then undergoes a rate-determining structural change. The second is the monomer-directed conversion (MDC) mechanism \(^\text{13}\), in which a monomeric peptide adopts a conformation that is identical to that found in the fibril, binds to another soluble monomer in a rate-determining step, converts that monomer to the conformation found in the fibril, and then dissociates, with both peptides rapidly added to the end of the growing fibril. The third is the nucleated polymerization (NP) mechanism \(^\text{14–16}\), which is characterized by the rate-limiting formation of a nucleus, followed by the rapid addition of monomers to the growing end of the nucleus. The fourth is the nucleated-conformational conversion (NCC) mechanism \(^\text{17}\), in which the formation of small amorphous aggregates precedes the rate-limiting formation of a critical nucleus and subsequent rapid growth of large fibrils. All of the above-mentioned mechanisms require that the system go through a nucleation event before fibrils can be formed. The nucleus in the MDC mechanism is a monomeric peptide whereas the nuclei in the TA, NP, and NCC mechanisms are oligomers.

The general picture of the fibril formation process that accompanies three of the four mechanisms described above - the TA, NP, and NCC mechanisms - is the following. Fibril formation is initiated when the native state of a protein is slightly destabilized, for example by changing pH, exposing structural elements on the resulting partially-folded intermediates which then begin
to associate intermolecularly rather than intramolecularly. If the protein monomer concentration is greater than some critical concentration, the partially-folded intermediates slowly associate via a series of energetically unfavorable steps, resulting in the formation of an oligomer after a defined period, called the lag time. This oligomer serves as the nucleus for the rapid growth and elongation of protofilaments, either through monomer addition at the protofilament tip or end-to-end association of short protofilaments. Eventually several protofilaments associate to form fibrils which then grow either by the addition of shorter fibrils or protofilaments. If the protein monomer concentration is less than the critical concentration, fibrillization can still take place by homogeneous (or heterogeneous) seeding in which case nucleation occurs on protein (or non-protein) species.

Most simulation studies to date of fibril-forming peptides by other investigators have been limited to the study of either isolated peptides or model amyloid fibrils that have already formed. These studies have employed high-resolution protein models, which are based on a realistic representation of protein geometry and a fairly faithful accounting for the energetics of every atom on the protein and on the solvent. Although there have been several attempts using high-resolution protein models to simulate the formation of fibrils from random coils, the systems considered did not contain enough peptides to mimic the nucleus that stabilizes the large fibrils that are observed in experiments. Given current computational capabilities, simpler models are required. This has been recognized by a few investigators who have combined intermediate-resolution protein models with the Go potentials to look at fibril formation. Such an approach has been taken by Jang et al., who studied the thermodynamics and kinetics of the assembly of four model \(\beta\)-sheet peptides into a tetrameric \(\beta\)-sheet complex, and by Ding et
al.\textsuperscript{44}, who studied the formation of a fibrillar double $\beta$-sheet structure containing eight model Src SH3 domain proteins. However, since the Go potential contains a built-in bias toward the native conformation, this approach is not suitable for the study of spontaneous fibril formation from random configurations.

We have developed an alternative approach which allows the treatment of large multi-chain systems while maintaining a fairly realistic description of protein dynamics without built-in bias toward any conformation\textsuperscript{45,46}. By combining our intermediate resolution protein model (described below) with discontinuous molecular dynamics simulation, we have been able to simulate the formation of fibrils by systems containing between 12 and 96 16-residue Ac-KA\textsubscript{14}K-NH\textsubscript{2} peptides starting from the random-coil state. This model, which was developed by Smith and Hall \textsuperscript{45,47,48} and later improved by us\textsuperscript{46}, represents each amino acid with four beads - three for the backbone and one for the sidechain. This model is designed to be used with discontinuous molecular dynamics (DMD)\textsuperscript{49–52}, an extremely fast alternative to traditional molecular dynamics, that is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials. Solvent is modeled implicitly by including hydrophobic interactions between non-polar sidechains. Backbone hydrogen bonding is modeled in explicit detail. Using this algorithm, we \textsuperscript{53} were able to sample much wider regions of conformational space, longer time scales, and larger systems than in traditional molecular dynamics. Since the simulations take only days on a workstation, we were able to conduct simulations at a wide variety of concentrations and temperatures, and to learn how peptide concentration and temperature affect the formation of various Ac-KA\textsubscript{14}K-NH\textsubscript{2} structures, including amorphous aggregates, $\alpha$-helices, $\beta$-sheets and fibrils. All simulations were performed in the canonical ensemble starting
from a random coil configuration equilibrated at a high temperature and then slowly cooled to
the temperature of interest so as to minimize kinetic trapping in local free energy minima.

In this paper, we investigate the kinetics of fibril formation of Ac-KA_{14}K-NH_{2} peptides as
a function of the peptide concentration and temperature. Simulations are conducted on systems
containing 48 model 16-residue peptides at a wide variety of concentrations and temperatures
using the same protein model as in our previous studies\textsuperscript{53}. All simulations are performed in the
canonical ensemble; constant temperature is achieved by implementing the Andersen thermostat
method\textsuperscript{54} in which united atoms are subjected to random collisions with ghost particles whose
velocities are chosen randomly from a Maxwell Boltzmann distribution centered at the system
temperature. During each simulation, the formation of different structures such as \(\alpha\)-helices,
aggregates, \(\beta\)-sheets or fibrils is monitored as a function of time. Key fibril-forming events that
are associated with the four proposed fibril forming mechanisms mentioned earlier are identified.

Two types of simulations are conducted: unseeded and seeded. In the unseeded simulations, the
peptides, which are initially in a random coil configuration, are equilibrated at a high temperature
and then quickly cooled to the temperature of interest. The lag time and rate of fibril formation
are monitored to study their dependence upon the peptide concentration and temperature. In the
seeded simulations, a previously created single fibril is immersed in a sea of denatured chains.

The model peptide chosen for study is the polyalanine-based peptide Ac-KA_{14}K-NH_{2}. We
focus on polyalanine-based peptides for three reasons. First, the small, uncharged, unbranched
nature of alanine residues is amenable to simulation with the intermediate-resolution protein
model that we developed previously\textsuperscript{45,47}. Second, polyalanine repeats have been implicated in
human pathologies, notably in the formation of anomalous filamentous intranuclear inclusions
in oculopharyngeal muscular dystrophy (OPMD) patients. Third, synthetic polyalanine-based peptides have been shown by Blondelle et al. to undergo a transition from \( \alpha \)-helical structures to \( \beta \)-sheet complexes \textit{in vitro}, mimicking the structural transition that is believed to be a prerequisite for fibril nucleation and growth. Blondelle et al. observed that the \( \alpha \)-helical structures were stabilized in part by intramolecular \( \alpha \)-helical bonds and the macromolecular \( \beta \)-sheet complexes were stabilized by hydrophobic inter-sheet interactions. Using circular dichroism, fourier-transform infrared spectroscopy, and reversed phase high performance liquid chromatography, they found that: (1) \( \beta \)-sheet complex formation increased with increasing temperature, exhibiting an S-shaped dependence on temperature with a critical temperature of 45°C at a peptide concentration of 1.8mM and an incubation time of 3h, and (2) \( \beta \)-sheet complex formation increased with increasing peptide concentration above a critical concentration of 1mM at 65°C.

Highlights of our results are the following. We observe that fibril formation of polyalanines in our simulations is not perfectly described by any of the proposed fibril-formation mechanisms available in the literature. It does, however, share certain features in common with three of the proposed mechanisms: templated assembly, nucleated polymerization, and nucleated conformational conversion. These common features are: (1) fibril formation is nucleation dependent, and (2) the lag time, which is the delay time before fibril formation commences, increases with increasing temperature. As suggested by the nucleated polymerization model, the lag time at high temperatures decreases more-or-less exponentially as a function of the peptide concentration. As suggested by the templated assembly model, the lag time at low temperatures decreases more-or-less linearly with increasing peptide concentration. As predicted by the nucleated confor-
mational conversion model, fibril formation involves small amorphous aggregate formation that precedes critical nucleus formation and subsequent rapid growth of large fibrils. Fibril growth in our simulations involves both $\beta$-sheet elongation, in which the fibril grows by adding individual peptides to the end of each $\beta$-sheet and lateral addition, in which the fibril grows by adding already-formed $\beta$-sheets to its side. The initial rate of fibril growth increases with increasing concentration and decreases with increasing temperature.

This paper is organized as follows. Section 2 describes the protein model used for this work, including the physical representation and the potential energy function, and the DMD simulation technique. Section 3 presents and discusses the results obtained from unseeded and seeded simulations. Section 4 provides a summary and a brief discussion of the work.

5.2 Method

5.2.1 Model peptide and forces

The model peptide has the sequence PH$_{14}$P, where H is a hydrophobic amino acid residue and P is a polar amino acid residue. This sequence was chosen to mimic Ac-KA$_{14}$K-NH$_2$ peptides which have been shown to form stable, soluble $\beta$-sheet complexes \textit{in vitro}.$^{56,57}$ The peptide is modeled at an intermediate level of resolution $^{45,47,48}$, based on a four-bead amino acid representation with realistic bond lengths and bond-angle constraints, and has the ability to interact both intra- and inter-molecularly via hydrogen bonding and hydrophobic interaction potentials. The geometry of the protein model is illustrated in Figure 5.1. Each amino acid residue is composed of four spheres, a three-sphere backbone comprised of united atom NH, C$_{\alpha}$H, and C=O, and a single-bead sidechain R (labeled N, C$_{\alpha}$, C, and R, respectively in the figure). All backbone
bond lengths and bond angles are fixed at their ideal values; the distance between consecutive C\(\alpha\)
atoms is fixed so as to maintain the inter-peptide bond in the trans configuration. The sidechains
are held in positions relative to the backbone so that all residues are L-isomers. Details of the
model including values for all parameters are given in our earlier papers \(^{45,46}\).

The solvent is modeled implicitly in the sense that its effect is factored into the energy func-
tion as a potential of mean force. All forces are modeled by either hard-sphere or square-well po-
tentials. The excluded volumes of the four beads are modeled using hard-sphere potentials with
realistic diameters. Covalent bonds are maintained between adjacent spheres along the backbone
by imposing hard sphere repulsions whenever the bond lengths attempt to move outside of the
range between \(l(1-\delta)\) and \(l(1+\delta)\) where \(l\) is the bond length and \(\delta\) is a tolerance which we set
equal to 2.375\%. Ideal backbone bond angles, C\(\alpha\)-C\(\alpha\) distances and residue L-isomerization are
achieved by imposing pseudobonds, as shown in Figure 5.1, which also fluctuate within a toler-
ance of 2.375\%. Interactions between hydrophobic sidechains are represented by a square-well
potential of depth \(\epsilon_{HP}\) and range 1.5\(\sigma_R\) where \(\sigma_R\) is the sidechain bead diameter. Hydrophobic
sidechains must be separated by at least three intervening residues in order to interact. Hydrogen
bonding between amide hydrogen atoms and carbonyl oxygen atoms on the same or neighbor-
ing chains are represented by a square-well attraction of strength \(\epsilon_{HB}\) between NH and C=O
united atoms whenever: (1) the virtual hydrogen and oxygen atoms (whose location can be cal-
culated at any time) are separated by 4.2\(\AA\) (the sum of the NH and C=O well widths), (2) the
nitrogen-hydrogen and carbon-oxygen vectors point towards each other within a fairly generous
tolerance, (3) neither the NH nor the C=O are already involved in a hydrogen bond with a differ-
ent partner, and (4) the NH and C=O are separated by at least three intervening residues along
the chain. To satisfy the second requirement, the separations between the four auxiliary pairs \( N_i-C_{a,j}, N_i-N_{j+1}, C_j-C_{a,i}, C_j-C_{i-1} \) surrounding the hydrogen bond in question are limited to certain distances that are chosen to maintain ideal hydrogen bond angles. This is accomplished by imposing square-shoulder interactions between the auxiliary pairs as suggested by Ding et al.\(^6^4\). Besides adding stability to the hydrogen bond, these interactions exact a penalty for breaking a hydrogen bond when any one of these auxiliary pairs moves inside the specified separation and thus distorts the hydrogen bond angle. For more details on the hydrogen bonding model used here, see the paper by Nguyen and Hall \(^4^6\). For simplicity, the ratio of the strength of a hydrophobic contact \( \epsilon_{HP} \), and the strength of a hydrogen bond \( \epsilon_{HB} \), \( R \equiv \epsilon_{HP}/\epsilon_{HB} \) is set equal to 1/10. Hydrogen bond strength and hydrophobic contact strength are independent of temperature, as has been assumed in previous simulation studies \(^6^5, ^4^7, ^4^8\).

5.2.2 Discontinuous molecular dynamics

Simulations are performed using the discontinuous molecular dynamics (DMD) simulation algorithm \(^4^9–^5^2\), which is an extremely fast alternative to traditional molecular dynamics and is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials. Unlike soft potentials such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide, enabling the exact (as opposed to numerical) solution of the collision dynamics. DMD simulations proceed in the following fashion. The initial positions of the beads on the model protein are assigned at random, but may not violate any of the size constraints or assigned bond lengths and angles. The initial velocities are chosen at random from a Maxwell-Boltzmann distribution at a fixed reduced temperature, \( T^* = k_B T/\epsilon_{HB} \),
where $k_B$ is Boltzmann’s constant, $T$ is the temperature, and $\epsilon_{HB}$ is the depth of the square-well hydrogen bond potential. The simulation proceeds according to the following schedule: identify the first event, move forward in time until that event occurs, calculate new velocities for the pair of beads involved in the event and calculate any changes in system energy resulting from hydrogen bond events or hydrophobic interactions, find the second event, and so on. Types of events include excluded volume events, bond events, and square-well hydrogen bond and hydrophobic interaction events. An excluded volume event occurs when the surfaces of two hard spheres collide and repel each other. A bond (or pseudobond) event occurs via a hard-sphere repulsion when two adjacent spheres attempt to move outside of their bond length. Square-well events include well-capture, well-bounce, and well-dissociation “collisions” when a sphere enters, attempts to leave, or leaves the square well of another sphere. For more details on DMD simulations with square-well potentials, see papers by Alder and Wainwright and Smith, Hall, and Freeman.

Simulations are performed in the canonical ensemble which means that the number of particles, volume, and temperature are held constant. Constant number of particles and volume are achieved by creating a virtual three-dimensional box for the simulation and allowing the model protein chains to move within that box. Periodic boundary conditions are used to eliminate artifacts due to simulation box walls. The dimensions of the box are chosen to ensure that a chain cannot interact with more than one image of any other chain. For this study, we use cubic boxes with sides ranging from 200 to 430 Å in length depending on the peptide concentration. Constant temperature is achieved by implementing the Andersen thermostat method as was used previously. With this procedure, all beads in the simulation are subject to random collisions.
with ghost particles. The post-event velocity of a bead colliding with a ghost particle is chosen randomly from a Maxwell-Boltzmann distribution at the simulation temperature. Time is measured in terms of the reduced time \( t^* \), which is defined as \( t/\sigma\sqrt{k_B T/m} \) with \( t \) the simulation time, and \( \sigma \) and \( m \) the average bead diameter and mass, respectively.

In the unseeded simulations, each system is started from a random coil configuration equilibrated at a high temperature, \( T^* = 0.25 \) and then quickly cooled down to the temperature of interest, \( T^* = 0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14, \) and \( 0.15 \). The number of ghost collisions is set at 1.0\% of the total number of collisions during a simulation. The resulting cooling rate is \( \Delta T^*/\Delta t^* = 0.04 \) during the first one million collisions (up to \( t^* = 0.1 \)), after which the system temperature remains constant. Simulations are conducted on systems containing 48 peptides at concentrations \( c = 1, 2.5, 3.75, 5, \) and 10mM. At each temperature and concentration, at least ten simulations are conducted.

In the seeded simulations, each of the ten systems initially contains fibrillar structures, which were obtained at the end of the ten \( c = 1.0 \) mM and \( T^* = 0.13 \) unseeded 48-peptide simulations described above. Forty eight denatured chains are added to the simulation box by randomly picking empty spaces in the simulation box and inserting the new denatured chains one-by-one. Each of the resulting 96-peptide systems at \( c = 2.0 \) mM is quickly heated to the temperature of interest, \( T^* = 0.14 \) and \( 0.15 \).

All systems were simulated for very long times. The simulations were stopped when the ensemble averages of the system’s total potential energy varied by no more than 2.5\% over the last three quarters of the simulation run. Simulations took between 40 hours at \( T^* = 0.08 \) and 160 hours at \( T^* = 0.15 \) on a single processor of an AMD Athlon MP 2200+ workstation.
The results presented in this paper are averages from at least ten simulations at each temperature and concentration with error bars taken from the standard deviations at each time. The structures are defined in the following way. If 12 intra-peptide α-helical hydrogen bonds (bonds between $N_{i+4}$ and $C_i$) are formed, the structure is an α-helix. If each peptide in a group of peptides has at least 7 inter-peptide β-hydrogen bonds to a particular neighbouring peptide in the group, this group as a β-sheet. If at least 2 β-sheet structures form inter-sheet hydrophobic interactions (at least 4 hydrophobic interactions per peptide per β-sheet) and the β-sheet structures are at an angle less than 35°, this as a fibril; otherwise, this and isolated β-sheets are classified as non-fibrillar β-sheet structures. If each peptide in a group of peptides has at least 2 inter-peptide hydrogen bonds or hydrophobic interactions with a neighbouring peptide in the same group, that group is an aggregate. Aggregates that are not fibrils or non-fibrillar β-sheets are amorphous aggregates. Single-peptide structures that are not α-helices or β-structures are random coils.

5.3 Results and Discussion

Since this paper builds upon our previous work on the fibril formation of peptides of the same sequence, it is useful to briefly review those results that are pertinent to the discussion here. We investigated how peptide concentration and temperature affect the formation of various Ac-KA$_{14}$K-NH$_2$ structures including α-helices, β-sheets and fibrils. Simulations were conducted on systems of 12, 24, 48, and 96 model 16-residue peptides at a wide variety of concentrations and temperatures by applying the discontinuous molecular dynamics simulation algorithm to our intermediate-resolution protein model. All simulations were performed in the canonical ensemble starting from a random coil configuration equilibrated at a high temperature and then
slowly cooled to the temperature of interest so as to minimize kinetic trapping in local free energy minima. Structural characteristics such as the peptide arrangement and packing within fibrils were examined and compared with those observed in experiments. We also studied the overall stability of fibrils by conducting simulations on already-formed fibrils over a wide range of temperatures to investigate the relative importance of hydrogen bonding and hydrophobic interactions on fibril stability. The stability of our fibrillar structures was evaluated by comparing the abilities of the system to maintain the fibrillar structures at various temperatures which were higher than the fibril formation temperature.

We were able to observe the formation of fibrils containing 12 to 96 polyalanine peptides starting from random coils in a relatively short period of time ranging between 40 to 160 hours on a single processor of an AMD Athlon MP 2200+ workstation. To our knowledge, these were the first simulations to span the whole process of fibril formation from the random coil state to the fibril state on such a large system. We found that there was a strong relationship between the formation of $\alpha$-helices, $\beta$-sheets, aggregates and fibrils and the environmental conditions such as temperature, concentration, and hydrophobic interaction strength. At low concentrations, $c < 1.0\text{mM}$, random-coil peptides assembled into $\alpha$-helices at low temperatures, and random coils at high temperatures, $T^* > 0.10$. At intermediate concentrations, $c = 1.0 - 2.5\text{mM}$, random-coil peptides assembled into $\alpha$-helices at low temperatures, $T^* = 0.08 - 0.10$, and large $\beta$-sheet structures at high temperatures, $T^* = 0.11 - 0.14$. At high concentrations, $c > 2.5\text{mM}$, random-coil peptides formed $\beta$-sheets over a wide range of temperatures, $T^* = 0.08 - 0.14$. These $\beta$-sheets assembled into fibrils above a critical temperature that decreased with concentration and exceeded the isolated peptide’s folding temperature ($T^* = 0.11$). At very high temperatures, $T^* \geq \ldots$
0.15, and all concentrations, the system was in a random-coil state. The critical concentration for fibril formation was between $c = 1.0\text{mM}$ and $c = 2.5\text{mM}$ at high temperatures $T^* = 0.12 - 0.14$. These results are in qualitative agreement with experimental results of Blondelle et al. on Ac-KA$_{14}$K-NH$_2$ peptides$^{56,57}$. The fibrils observed in our simulations mimicked the structural characteristics observed in experiments in that most of the peptides in our fibrils were arranged in an in-register parallel orientation, with intra-sheet and inter-sheet distances similar to those observed in experiments, and contained about six multi-peptide $\beta$-sheets. Finally, we found that when the strength of the hydrophobic interaction between non-polar sidechains relative to the strength of the hydrogen bonds was increased from $R = 1/10$ to $R = 1/6$, the system formed amorphous rather than fibrillar aggregates. We also identified key fibril-forming events. Since simulations were conducted by slowly cooling the system down to the temperature of interest, analysis of the temperature-dependence of the kinetics of fibril formation was not appropriate.

### 5.3.1 Fibril formation is preceded by the formation of amorphous aggregates

We find that fibril formation is a conformational conversion process in which the appearance of amorphous aggregates precedes $\beta$-sheet formation and then nucleus formation. This can be seen in Figure 5.2 which shows snapshots of the fibrillization process, which were taken at various reduced times, $t^*$, for the 48-peptide simulation at $T^* = 0.14$ and a peptide concentration of $c = 10\text{mM}$. The system was initially equilibrated at a high temperature, $T^* = 0.25$ and then quickly cooled down to $T^* = 0.14$ within 0.1 reduced time units. In the snapshots, all hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets once the fibril is formed. Starting in random
coil conformations at reduced time $t^*=0$, the chains begin to form small amorphous aggregates almost immediately. These amorphous aggregates have grown by $t^*=10.0$ and have collapsed into one big amorphous aggregate by $t^*=12.0$. This big amorphous aggregate is not stable and thus disperses by $t^*=14.6$ into smaller aggregates, one of which is a two-peptide $\beta$-sheet (i.e., the purple sheet at the middle right side of the box). By $t^*=36.1$, the other amorphous aggregates have converted into three three-peptide $\beta$-sheets (i.e., the light blue sheet in the middle of the box, the white sheet behind the light blue sheet, and the dark blue sheet in the upper right corner of the box); in addition, the purple sheet has grown into a six-peptide $\beta$-sheet. By $t^*=40.2$, the light blue sheet has associated with the white sheet creating a two-sheet fibrillar structure. By $t^*=49.7$, this two-sheet fibrillar structure has been joined by the dark blue sheet and the purple sheet. However, the blue sheet is at an oblique angle with the other sheet. By $t^*=92.7$ the blue sheet has re-aligned itself so that all of its peptides are parallel with those in the fibrillar aggregate forming a four-sheet fibril, which itself has grown by adding peptides to the ends of $\beta$-sheets. Even after a long equilibration time, $t^*=205.9$, the fibrillar structure remains stable and grows by adding more peptides to the ends of the $\beta$-sheets. The fibrillization process just described exhibits a sequence of events that is typical of many of our simulations at different peptide concentrations and temperatures. In essence, at the beginning of the fibrillization process, the system of denatured peptides stays in a lag phase during which some amorphous aggregates form. These aggregates then convert themselves into small $\beta$-sheets containing aligned $\beta$-strands. Once these $\beta$-sheet structures attain a certain size, they come together and align one-by-one, creating a fibril. The fibril grows by adding peptides to the end of each $\beta$-sheet.

Quantitative analysis of our data confirms that our simulations exhibit the conformational
conversion from amorphous aggregates to small fibrillar structures with subsequent rapid growth of large fibrils suggested by Serio et al.\textsuperscript{17}. This can be seen in Figure 5.3, which plots the percentage of peptides in aggregates of all types (amorphous aggregates, fibrils and non-fibrillar $\beta$-sheets), $\beta$-sheets (fibrillar and non-fibrillar) and fibrillar structures, the average number of peptides per $\beta$-sheet per fibril and the average number of $\beta$-sheets per fibril over time $t^*$. The data here is taken from the simulation at $T^* = 0.14$ and $c = 10\text{mM}$ shown in Figure 5.2. As indicated in Figure 5.3(a), aggregates are formed instantaneously. There is a delay time of 20 reduced time units before some of the aggregates convert into a $\beta$-sheets and a lag time of 32 reduced time units before fibrils start to appear. As indicated in Figure 5.3(b), the early fibrils are relatively small, containing two $\beta$-sheets each consisting of two peptides. The number of $\beta$-sheets per fibril increases over time indicating that the fibril grows in part by adding already-formed $\beta$-sheets to its side. In addition, the number of peptides per $\beta$-sheet increases gradually with time, indicating that the fibril also grows by adding peptides to the end of each $\beta$-sheet, thereby lengthening along the fibril axis. Even after the fibril reaches its final size of four $\beta$-sheets, the number of peptides per $\beta$-sheet per fibril continues to increase from 6 to 9 peptides.

We can offer an explanation for the presence of amorphous as opposed to ordered aggregates such as $\beta$-sheets and small fibrils, both in our model system and in experimental systems, based on an analysis of their energetic properties. For a small ordered aggregate, few residues are completely buried in the core of the structure, so most of the residues are on the surface. Though the ordered nature of the peptide chains in a $\beta$-sheet or fibril yields a dense array of hydrogen bonds and hydrophobic contacts in its core, if the core is small, there are a significant number of exposed hydrophobic side chains and unsatisfied hydrogen bond donors and acceptors on the
surface. Amorphous aggregates, on the other hand, tend to contain a web of hydrogen bonds and hydrophobic contacts that is only slightly more dense at its center than on its surface. If just a few peptide chains aggregate, an amorphous structure may contain more hydrogen bonds and hydrophobic contacts than an ordered structure, and hence be energetically more favorable. As the aggregate size increases, formation of ordered structures is more energetically favorable.

5.3.2 The rate of fibril formation is a function of peptide concentration and temperature

Simulations are conducted on unseeded 48-peptide systems initially in a random-coil state at concentrations $c = 2.5, 3.75, 5$, and $10$ mM, which span the range of concentrations at which fibrils can be formed, according to our previous study. At each concentration, simulations are performed at constant temperatures of $T^* = 0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14, 0.15$, which range from a temperature that is well below the folding temperature for a single $\alpha$-helix ($T^* = 0.11$) to a temperature ($T^* = 0.15$) that is so high that peptides cannot form or maintain hydrogen bonds or hydrophobic interactions, i.e. they are in random-coil configurations. The results here are average values from at least ten simulation at each temperature and concentration. The errors bars are standard deviations from the average values at each time.

The rate of fibril formation increases with increasing concentration as can be seen in Figure 5.4, which plots the percentage of peptides in aggregates of all types (amorphous aggregates, fibrils and non-fibrillar $\beta$-sheets), $\beta$-sheets (fibrillar and non-fibrillar), and fibrillar structures as a function of reduced time $t^*$ for the 48-peptide system at a constant temperature of $T^*=0.14$ and different peptide concentrations, $c = 2.5, 3.75, 5.0$, and $10.0$ mM. The rates of forming aggregates
and β-sheets also increase with increasing concentration. This figure also indicates that the formation of aggregates precedes the formation of β-sheets, which itself precedes the formation of fibrils. In addition, the percentage of peptides in β-sheets is less than the percentage of peptides in aggregates, indicating that not all of the aggregated peptides convert into β-sheets. Likewise, the percentage of peptides in fibrils is less than the percentage of peptides in β-sheets meaning that not all of the β-sheets become fibrils.

The rate of fibril formation depends on the temperature as can be seen in Figure 5.5, which plots the percentage of peptides in aggregates of all types (amorphous aggregates, fibrils and non-fibrillar β-sheets), β-sheets (fibrillar and non-fibrillar), and fibrillar structures as a function of time \( t^* \) for the 48-peptide system at a constant concentration of \( c = 10.0 \text{mM} \) and different temperatures, \( T^* = 0.13 \), and 0.14. The rate of forming aggregates and β-sheets decreases with increasing temperatures. In terms of forming β-sheets, after the system at \( T^* = 0.13 \) reaches a plateau at approximately 85%, the system at \( T^* = 0.14 \) continues to form more β-sheets, reaching a plateau at 85% by \( t^* = 200 \). Although the initial rate of forming fibrils at \( T^* = 0.13 \) is greater than that at \( T^* = 0.14 \), the rate of fibril formation at \( T^* = 0.13 \) later in the simulation is less than that at \( T^* = 0.14 \). After the system at \( T^* = 0.13 \) reaches a plateau at approximately 40% at \( t^* = 100 \), the system at \( T^* = 0.14 \) continues to form more fibrils. The reason that the system at \( T^* = 0.13 \) stops growing fibrils after the early stages in the simulation is that it is more likely to be kinetically trapped than the system at \( T^* = 0.14 \). At \( T^* = 0.13 \), β-sheets are more likely to collapse onto one another at an oblique angle larger than 35°, thus not satisfying our criteria for classifying an aggregate as a fibril.
5.3.3 Fibril formation lag time as a function of peptide concentration and temperature

The lag time for the formation of fibrils observed in our simulations decreases with increasing peptide concentration and increases with increasing temperature as indicated by Figure 5.6, which plots the lag time $t_{\text{lag}}^*$ versus the concentration for the 48-peptide system at $T^*=0.13$ and $T^*=0.14$. At $T^*=0.14$, the lag time decreases exponentially with increasing peptide concentration. However, at $T^*=0.13$, the decrease in the lag time with peptide concentration does not appear to be an exponential decay; instead, the lag time decreases only slightly with increasing peptide concentration. The rates of decrease in our lag time with peptide concentration are not completely consistent with any of the four proposed mechanisms described in the literature. Although our lag time results at $T^*=0.14$ are consistent with the NP model, which predicts that $t_{\text{lag}}^* \propto \exp(-c)$, our lag time results at $T^*=0.13$ are more consistent with the TA model, which predicts that $t_{\text{lag}}^* \propto c$. Interestingly enough, our lag time results at all temperatures are not consistent with the NCC model of Serio et al., who observed that the lag time is relatively insensitive to the change in peptide concentration; they found, for example, that the lag time over a 500-fold range of concentration decreases by only less than 10-fold.

5.3.4 Fibril growth proceeds by both $\beta$-sheet elongation and lateral addition

There are two mechanisms of fibril growth in our simulations: elongation and lateral addition. In $\beta$-sheet elongation, the fibril grows by adding individual peptides to the end of each $\beta$-sheet. In lateral addition, the fibril grows by adding already-formed $\beta$-sheets to its side. Both mechanisms can be seen in the snapshots in Figure 5.2 and in the data for $c = 10\text{mM}$ and $T^* = $
0.14 on the average number of peptides per β-sheet in a fibril and the average number of β-sheets per fibril in Figure 5.3(b). These two growth mechanisms are observed in all of our simulations at different conditions as illustrated in Figure 5.7, which plots the average number of peptides per β-sheet in a fibril and the average number of β-sheets per fibril versus reduced time, $t^*$, for the 48-peptide system at a constant temperature of $T^*=0.14$ and different peptide concentrations, $c=2.5, 3.75, 5.0, \text{ and } 10.0\text{mM}$. Each data point is an average value from at least ten simulations. Figure 5.7 shows that there is a more-or-less gradual increase in the average number of peptides per β-sheet in a fibril as a function of time, increasing from two to nine peptides per β-sheet in a fibril. Figure 5.7 also shows that there is a gradual increase in the average number of β-sheets per fibril as a function of time, increasing from two to five β-sheets per fibril. Towards the end of each simulation, the fibrils contain between three and five β-sheets, each with six to nine peptides. This was also observed in our previous simulation study \cite{53} of the slow-cooling formation of fibrils in systems containing 12 to 96 peptides at $c=5\text{mM}$, $R=1/10$, and $T^*=0.13$. Once the fibrillar structure reaches its critical β-sheet number, monomeric peptides tend to attach to it rather than creating an isolated β-sheet. These two growth mechanisms are similar to those observed in experiments by Green et al.\cite{68}, who found two distinct phases in human amylin (hA) fibrillogenesis in which lateral growth of oligomers is followed by longitudinal growth into mature fibrils.

There is an energetic explanation for the tendency of the fibrils to grow very long along the fibril axis rather than to grow laterally by including more and more β-sheets. We can imagine growth of the complex as occurring either by addition of a peptide that extends a sheet or by addition of a peptide that creates a new sheet. This is illustrated in Figure 5.8 which shows the
numbers of hydrogen bonds (HB) and hydrophobic interactions (HP) between an inner peptide, labeled “P”, and the adjacent peptides in the structure. Positions A and B represent possible extensions of the fibril core. A peptide in position A, which represents β-sheet elongation, would form 15 hydrogen bonds and 27 hydrophobic interactions (14 intra-sheet and 13 inter-sheet) with the existing fibril scaffold while a peptide at position B, which represents β-sheet creation, would form only 26 hydrophobic interactions. Therefore, there is an energetic preference for β-sheet elongation (position A) as opposed to β-sheet creation (position B). This energetic preference would help to explain the asymmetric fibril growth that is seen in nature, where fibrils are composed of four to six β-sheets and each β-sheet is made up of hundreds or thousands of protein chains.

5.3.5 Fibril formation is nucleation dependent

Fibril formation involves a nucleation event as suggested by our seeded simulations in which previously created fibrils are immersed in a sea of denatured chains. The results from these seeded simulations at $T^*=0.15$ are shown in Figure 5.9, which plots the average number of peptides per β-sheet per fibril and the average number of β-sheets per fibril over time $t^*$. This figure indicates that the average number of β-sheets per fibril remains constant at four while the average number of peptides per β-sheet per fibril increases from approximately five to almost seven over time. This is consistent with data (not shown) indicating that there are no new fibrils formed in the system. Denatured random-coil peptides form fibrils not by creating new ones but by attaching themselves onto the seeded fibril. In other words, the seeded fibril grows at $T^*=0.15$ by adding monomeric denatured peptides to the ends of its β-sheets. This is in
contrast to the constant-temperature unseeded simulation results presented here and the results from our previous slow-cooling unseeded simulation paper\textsuperscript{53} in which fibril formation occurs only at temperatures up to $T^*=0.14$. In all unseeded simulations, the kinetic energy is too high at at $T^*=0.15$ to form and maintain any hydrogen bonds or hydrophobic interactions; therefore, peptides at all concentrations considered are random coils (data not shown). Comparison of the results from seeded and unseeded simulations at $T^*=0.15$ suggests that fibril formation is a nucleation-dependent process, which is similarly observed in various experimental studies\textsuperscript{3,73,17,74}.

Fibril formation in our simulations can bypass the slow nucleation step in the presence of preformed nucleus or seed. This is shown in Figure 5.10, which plots the percentage of the 48 random-coil peptides inserted into the seeded systems that attach to the fibrils (resulting in fibril growth) as a function of the reduced time, $t^*$, at $c=2.0mM$ and $T^*=0.14$. The percentage of peptides that are in fibrils during the unseeded simulations at the same condition are also plotted for comparison. The lag time for fibril formation from the unseeded simulations is about 135 reduced time units compared to a lag time of zero for fibril formation in the unseeded simulations. This indicates that in the presence of a nucleus, random coils readily attach to the fibrils without going through the nucleation step after a long lag time.

5.4 Conclusions

Computer simulations offer unique opportunities to observe and analyze molecular-level events in fibril formation that are difficult or impossible to observe experimentally. In simulating large multi-peptide systems using an intermediate resolution protein model in conjunction
with the discontinuous molecular dynamics, we have been able to examine the kinetics of the fibrillization process of polyalanines to discern the molecular-level mechanisms responsible for nucleation and fibril growth at a variety of conditions. Two types of simulations were conducted: unseeded simulations and seeded simulations. In unseeded simulations, the initial systems contained random-coil peptides. In the seeded simulations, already-formed fibrils were immersed in a sea of denatured peptides.

The ability of a system to form fibrils at high temperatures depends upon whether there is a seeded structure present in the system. In the unseeded simulations, fibril formation occurs only at temperatures up to $T^* = 0.14$; at $T^* = 0.15$ and higher temperatures, peptides at all concentrations considered are random coils. However, in the seeded simulations at $T^* = 0.15$, random-coil peptides attach themselves to the seeded fibrils, indicating that in the presence of seeds or nuclei, fibrils can form at even higher temperature. At $T^* = 0.14$, fibril formation occurs quickly in the presence of nucleus compared to the long lag times that are seen in the unseeded simulations at the same temperatures. These results indicate that fibril formation is nucleation dependent, which is similarly observed in experiments $^{3,73,17,74}$.

In the unseeded simulations, there is a lag time before fibril formation commences; the lag time depends upon the temperature and peptide concentration. At high temperatures, the lag time decreases more-or-less exponentially as a function of concentration as suggested by the nucleated polymerization model $^{14-16}$. At low temperatures, the lag time decreases more-or-less linearly with increasing peptide concentration as suggested by the templated assembly model $^{11,12}$. In addition, fibril formation is preceded by the appearance of amorphous aggregates and then $\beta$-sheets, which is similar to the key features of the nucleated conformational conversion model.
proposed by Serio et al.\textsuperscript{17}. They observed that fibril formation by yeast prion protein Sup35 involves the formation of small amorphous aggregates containing 20 to 80 protein monomers which precedes critical nucleus formation and subsequent rapid growth of large fibrils. Fibril growth in our simulations involves both $\beta$-sheet elongation, in which the fibril grows by adding individual peptides to the end of each $\beta$-sheet, and lateral addition, in which the fibril grows by adding already-formed $\beta$-sheets to its side. Moreover, the initial rate of fibril formation increases with increasing concentration and decreases with increasing temperatures.

Despite the similarities between selected features observed in our simulations and the behavior predicted by the templated assembly, nucleated polymerization, and nucleated conformational conversion models, none of these gives a fully satisfactory description of the simulation kinetics of fibril formation by polyalanine peptides.

\section*{5.5 Acknowledgments}

The authors are grateful to Sylvie Blondelle, Ken Dill, Jeffery Kelly, and Dan Kirschner for helpful discussions. This work was supported by the National Institutes of Health under grant number GM-56766 and the National Science Foundation under grant number CTS-9704044.
5.6 References


[35] K. Kuwata et al., “NMR-detected hydrogen exchange and molecular dynamics simulations


5.7 Figures

5.1 Covalent bonds are shown with narrow black lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_{\alpha}$ distances, and residue L-isomerization. Note that the beads are not shown full size for ease of viewing. ................................................................. 224

5.2 Snapshots of the 48-peptide system at various times during a simulation at $c=10\text{mM}$ and $T^{*}=0.14$ that results in the formation of a fibril. All hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets once the fibril is formed. ................................................................. 225

5.3 (a) The percentage of peptides in aggregates (of any type), $\beta$-sheets, and fibril structures and (b) the average number peptides per $\beta$-sheet in a fibril and the average number of $\beta$-sheets per fibril versus reduced time, $t^{*}$, for the 48-peptide system at $c=10\text{mM}$ and $T^{*}=0.14$. ................................................................. 226

5.4 The percentage of peptides in aggregates (of any type), $\beta$-sheets, and fibril structures versus reduced time, $t^{*}$, for the 48-peptide system at a constant temperature of $T^{*}=0.14$ and different peptide concentrations $c=2.5\text{mM}$ (filled circles), $c=3.75\text{mM}$ (open squares), $c=5.0\text{mM}$ (filled diamonds), and $c=10.0\text{mM}$ (open triangles). ................................................................. 227
5.5 The percentage of peptides in aggregates (of any type), $\beta$-sheets, and fibrill structures versus reduced time, $t^*$, for the 48-peptide system at a constant concentration of $c=10.0\text{mM}$ and different temperatures $T^*=0.13$ (filled circles) and $T^*=0.14$ (open squares).

5.6 The lag time for fibril formation versus the concentration for the 48-peptide system at $T^*=0.13$ and $T^*=0.14$.

5.7 The average number of peptides per $\beta$-sheet in a fibril and the average number of $\beta$-sheets per fibril versus reduced time, $t^*$, for the 48-peptide system at a constant temperature of $T^*=0.14$ and different peptide concentrations $c=2.5\text{mM}$ (filled circles), $c=3.75\text{mM}$ (open squares), $c=5.0\text{mM}$ (filled diamonds), and $c=10.0\text{mM}$ (open triangles).

5.8 Schematic showing peptide locations within the fibril indicating the types of interactions (HP=hydrophobic interaction, HB=hydrogen bond) between an inner peptide, labeled “P”, and the adjacent peptides in the structure. Positions A and B represent possible extensions of the core of the fibril.

5.9 The average number of peptides per $\beta$-sheet in a fibril and the average number of $\beta$-sheets per fibril versus reduced time, $t^*$, from the seeded simulations of the 96-peptide system at $c=2.0\text{mM}$ and $T^*=0.15$.

5.10 The percentage of peptides in fibril structures versus reduced time, $t^*$, from the seeded and unseeded simulations of systems at $c=2.0\text{mM}$ and $T^*=0.14$. 
Figure 5.1: Covalent bonds are shown with narrow black lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_\alpha$ distances, and residue L-isomerization. Note that the beads are not shown full size for ease of viewing.
Figure 5.2: Snapshots of the 48-peptide system at various times during a simulation at $c=10\text{mM}$ and $T^*=0.14$ that results in the formation of a fibril. All hydrophobic sidechains are red; backbone atoms of different peptides have different colors, assigned so that it will be easy to distinguish the various sheets once the fibril is formed.
Figure 5.3: (a) The percentage of peptides in aggregates (of any type), \( \beta \)-sheets, and fibril structures and (b) the average number peptides per \( \beta \)-sheet in a fibril and the average number of \( \beta \)-sheets per fibril versus reduced time, \( t^* \), for the 48-peptide system at \( c=10 \text{mM} \) and \( T^*=0.14 \).
Figure 5.4: The percentage of peptides in aggregates (of any type), β-sheets, and fibril structures versus reduced time, $t^*$, for the 48-peptide system at a constant temperature of $T^* = 0.14$ and different peptide concentrations $c = 2.5$ mM (filled circles), $c = 3.75$ mM (open squares), $c = 5.0$ mM (filled diamonds), and $c = 10.0$ mM (open triangles).
Figure 5.5: The percentage of peptides in aggregates (of any type), β-sheets, and fibril structures versus reduced time, \( t^* \), for the 48-peptide system at a constant concentration of \( c=10.0\,\text{mM} \) and different temperatures \( T^*=0.13 \) (filled circles) and \( T^*=0.14 \) (open squares).
Figure 5.6: The lag time for fibril formation versus the concentration for the 48-peptide system at $T^*=0.13$ and $T^*=0.14$. 
Figure 5.7: The average number of peptides per β-sheet in a fibril and the average number of β-sheets per fibril versus reduced time, $t^*$, for the 48-peptide system at a constant temperature of $T^*=0.14$ and different peptide concentrations $c=2.5\text{mM}$ (filled circles), $c=3.75\text{mM}$ (open squares), $c=5.0\text{mM}$ (filled diamonds), and $c=10.0\text{mM}$ (open triangles).
Figure 5.8: Schematic showing peptide locations within the fibril indicating the types of interactions (HP=hydrophobic interaction, HB=hydrogen bond) between an inner peptide, labeled “P”, and the adjacent peptides in the structure. Positions A and B represent possible extensions of the core of the fibril.
Figure 5.9: The average number of peptides per β-sheet in a fibril and the average number of β-sheets per fibril versus reduced time, $t^*$, from the seeded simulations of the 96-peptide system at $c=2.0\text{mM}$ and $T^*=0.15$. 
Figure 5.10: The percentage of peptides in fibril structures versus reduced time, $t^*$, from the seeded and unseeded simulations of systems at $c=2.0\text{mM}$ and $T^*=0.14$. 
CHAPTER 6

THERMODYNAMICS OF FIBRIL FORMATION

BY POLYALANINE PEPTIDES

6.1 Introduction

The self-assembly of normally soluble proteins into stable, insoluble rodlike ordered aggregates, known as amyloid fibrils, is a cause or associated symptom of numerous neurodegenerative conditions, including Alzheimer’s and Parkinson’s diseases, the prion diseases and diabetic neuropathies. More than twenty biochemically amyloidogenic distinct proteins that are known to form in vivo fibrils which are structurally similar to one another. More than twenty biochemically amyloidogenic distinct proteins that are known to form in vivo fibrils which are structurally similar to one another. Moreover, a whole host of designed proteins not associated with amyloid diseases have been found to assemble in vitro into fibrils that are more or less indistinguishable from those found in neurodegenerative conditions. These findings imply that the fibril structure and fibrillization process are relatively independent of the sequence studied. Therefore, any simple peptide, say polyalanine which has one of the most simple amino acids and thus is easy to model, can be considered to be as good a representative of amyloidogenic proteins as any for a computational investigation of the basic physics of fibril formation. In this paper, we conduct equilibrium simulations on
96-peptide systems via the replica-exchange simulation method in an attempt to help understand
the thermodynamics of fibril formation with the ultimate goal of mapping out a phase diagram
in the temperature-concentration plane delineating the regions where random coils, \( \alpha \)-helices,
\( \beta \)-sheets, fibrils, and amorphous aggregates are stable.

Most simulation studies to date of fibril-forming peptides by other investigators have been
limited to the study of either isolated peptides \(^{14-22}\) or model amyloid fibrils that have already
formed \(^{23-30}\). These studies have employed high-resolution protein models, which are based on
a realistic representation of protein geometry and a fairly faithful accounting for the energetics
of every atom on the protein and on the solvent. Although there have been several attempts
\(^{31-35}\) using high-resolution protein models to simulate the formation of fibrils from random coils, the
systems considered do not contain enough peptides to mimic the nucleus that stabilizes the large
fibrils observed in experiments. Given current computational capabilities, simpler models are
required. This has been recognized by a few investigators who have combined intermediate-
resolution protein models with Go potentials to look at fibril formation. Such an approach has
been taken by Jang \textit{et al.}\(^{36,37}\), who studied the thermodynamics and kinetics of the assembly
of four model \( \beta \)-sheet peptides into a tetrameric \( \beta \)-sheet complex, and by Ding \textit{et al.}\(^{38}\), who
studied the formation of a fibrillar double \( \beta \)-sheet structure containing eight model Src SH3
domain proteins. However, since the Go potential contains a built-in bias toward the native
conformation, this approach is not suitable for the study of spontaneous fibril formation from
random configurations.

An alternative approach, which allows the treatment of large multi-chain systems while
maintaining a fairly realistic description of protein dynamics without built-in bias toward any
conformation, is that of Nguyen and Hall \textsuperscript{39}. By combining our intermediate resolution protein model (described below) with discontinuous molecular dynamics simulation, we were able to simulate the formation of fibrils by systems containing between 12 and 96 16-residue Ac-KA\textsubscript{14}K-NH\textsubscript{2} peptides starting from the random-coil state. Their model, which was developed by Smith and Hall \textsuperscript{40–42} and later improved by Nguyen and Hall \textsuperscript{43}, represents each amino acid with four beads - three for the backbone and one for the sidechain. This model is designed to be used with discontinuous molecular dynamics (DMD) \textsuperscript{44–47}, an extremely fast alternative to traditional molecular dynamics, that is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials. Solvent is modeled implicitly by including hydrophobic interactions between non-polar sidechains. Backbone hydrogen bonding is modeled in explicit detail. Using this algorithm, Nguyen and Hall \textsuperscript{39} were able to sample much wider regions of conformational space, longer time scales, and larger systems than in traditional molecular dynamics. All simulations were performed in the canonical ensemble starting from a random coil configuration equilibrated at a high temperature and then slowly cooled to the temperature of interest. Since the runs took only days on a workstation, they were able to conduct simulations at a wide variety of concentrations and temperatures, and to learn how peptide concentration and temperature affect the formation of various Ac-KA\textsubscript{14}K-NH\textsubscript{2} structures including amorphous aggregates, \(\alpha\)-helices, \(\beta\)-sheets and fibrils. Although kinetic trapping in local free energy minima was minimized by slowly cooling a system that was initially at a high temperature down to temperature of interest, they could never be certain if the system had reached equilibrium or gotten stuck in a metastable state.

In this paper, we perform equilibrium simulations on 96-peptide systems over a very wide
range of temperatures and peptide concentrations by using the replica exchange simulation method as originally formulated by Sugita and Okamoto for molecular dynamics simulations of protein folding. In this method, a number of replicas of the system are created and simulated at a spectrum of temperatures usually on a system of parallel computers. At set time intervals, replicas whose temperatures are nearest neighbors along the temperature spectrum are exchanged, provided that a Metropolis criterion is satisfied. This procedure is repeated until all of the systems at different temperatures reach equilibrium. Once at equilibrium, the data collection phase begins in which the probability of being in various energy levels and states are stored for use in a weighted histogram calculation of the conformational free energy of the various peptide structures. This simulation procedure is discussed in greater detail in the Methods section. The phase transitions of the systems at different concentrations are determined by calculating thermodynamic averages for the radius of gyration $R_g$, the specific heat $C_V$, and the internal energy $E$. The results are summarized in a phase diagram in the temperature-concentration plane.

The model polyalanine peptide chosen for study is the peptide Ac-KA$_{14}$K-NH$_2$. We focus on polyalanine-based peptides for three reasons. First, the small, uncharged, unbranched nature of alanine residues is amenable to simulation with the intermediate-resolution protein model that we developed previously. Second, polyalanine repeats have been implicated in human pathologies; in particular, they are responsible for the formation of anomalous filamentous intranuclear inclusions in oculopharyngeal muscular dystrophy patients. Third, synthetic polyalanine-based peptides have been shown by Blondelle et al. to undergo a transition from $\alpha$-helical structures to $\beta$-sheet complexes in vitro, mimicking the structural transition believed to be a prerequisite for fibril nucleation and growth. Blondelle et al. ob-
served that the α-helical structures were stabilized in part by intramolecular α-helical bonds and that the macromolecular β-sheet complex was stabilized by hydrophobic inter-sheet interactions. Using circular dichroism, fourier-transform infrared spectroscopy, and reversed phase high performance liquid chromatography, they found that: (1) β-sheet complex formation increased with increasing temperature, exhibiting an S-shaped dependence on temperature with a critical temperature of 45°C at a peptide concentration of 1.8mM and an incubation time of 3h, and (2) β-sheet complex formation increased with increasing peptide concentration above a critical concentration of 1mM at 65°C.

Highlights of our simulation results are the following. We find that there are four distinctive single-phase regions at which the following structures are stable: α-helices, fibrils, β-sheets and random coils and four different two-phase regions: random coils / unfibrillar β-sheets, random coils / fibrils, fibrils / unfibrillar β-sheets, and α-helices / unfibrillar β-sheets. The α-helical region is at low temperature and low concentration. The unfibrillar β-sheet region is at intermediate temperatures and expands to lower temperatures as concentration is increased. The fibril region is mostly at intermediate temperatures at intermediate concentrations and expands to lower and higher temperatures as the peptide concentration is increased. The random-coil region is at high temperatures and all concentration and shifts to higher temperatures as the concentration is increased.

This paper is organized as follows. Section 2 describes the methods used in this work, including the protein’s physical representation, its potential energy function, the DMD simulation technique, the replica-exchange method, and the weighted histogram method. Section 3 presents the results obtained from simulations at various conditions. Section 4 contains a summary and
discussion of our findings and a brief description of our work on fibril formation in multi-peptide systems.

6.2 Method

6.2.1 Model Peptide and Forces

The model peptide is 16 residues long with the sequence PH$\text{14}$P, where H stands for a hydrophobic amino acid residue and P stands for a polar amino acid residue. This sequence was chosen to approximate Ac-KA$\text{14}$K-NH$_2$ peptides which have been shown by Blondelle et al.$^{52,53}$ to form stable, soluble β-sheet complexes. The peptide is represented at an intermediate level of resolution using a model introduced by Smith and Hall.$^{40-42}$ Details of the model including values for all parameters are given in earlier papers. The model is based on a four-bead amino acid representation with realistic bond lengths and bond-angle constraints and has the ability to interact both intra- and inter-molecularly via hydrogen bonding and hydrophobic interaction potentials. The geometry of the protein model is illustrated in Figure 6.1. Each amino acid residue is composed of four spheres, a three-sphere backbone comprised of united atom NH, C$_\alpha$H, and C=O, and a single bead sidechain R (these are labeled N, C$_\alpha$, C, and R, respectively in the figure). All backbone bond lengths and bond angles are fixed at their ideal values; the distance between consecutive C$_\alpha$ atoms is fixed so as to maintain the interpeptide bond in the trans configuration. The sidechains are held in positions relative to the backbone such that all residues are L-isomers. Details of the model including values for all parameters are given in our earlier papers.$^{40,43}$

The solvent is modeled implicitly in the sense that its effect is factored into the energy func-
tion as a potential of mean forces. All forces are modeled by either hard-sphere or square-well potentials. The excluded volumes of the four united atoms are modeled using hard-sphere potentials with realistic diameters. Covalent bonds are maintained between adjacent spheres along the backbone by imposing hard sphere repulsions whenever the bond lengths attempt to move outside of the range between \( l(1-\delta) \) and \( l(1+\delta) \) where \( l \) is the bond length and \( \delta \) is a tolerance which we set equal to 2.375\%. Ideal backbone bond angles, \( C_{\alpha}-C_{\alpha} \) distances and residue L-isomerization are achieved by imposing pseudobonds, as shown in Figure 6.1, which also fluctuate within a tolerance of 2.375\%. Interactions between hydrophobic sidechains are represented by a square-well potential of depth \( \epsilon_{HP} \) and range \( 1.5\sigma_R \) where \( \sigma_R \) is the sidechain diameter. Hydrophobic sidechains must be separated by at least three intervening residues in order to interact. Hydrogen bonding between amide hydrogen atoms and carbonyl oxygen atoms on the same or neighboring chains are represented by a square-well attraction of strength \( \epsilon_{HB} \) between NH and C=O united atoms whenever: (1) the virtual hydrogen and oxygen atoms (whose location can be calculated at any time) are separated by 4.2\AA\ (the sum of the NH and C=O well widths), (2) the nitrogen-hydrogen and carbon-oxygen vectors point towards each other within a fairly generous tolerance, (3) neither the NH nor the C=O are already involved in a hydrogen bond with a different partner, and (4) the NH and C=O are separated by at least three intervening residues along the chain. To satisfy the second requirement, the separations between the four auxiliary pairs, \( N_i-C_{\alpha,j}, N_i-N_{j+1}, C_j-C_{\alpha,i}, C_j-C_{i-1} \), surrounding the hydrogen bond in question are limited to certain distances that are chosen to maintain ideal hydrogen bond angles. This is accomplished by imposing square-shoulder interactions between the auxiliary pairs. Besides adding stability to the hydrogen bond, these interactions exact a penalty for breaking a hydrogen bond when any
one of these auxiliary pairs moves inside the specified separation and thus distorts the hydrogen bond angle. For more details on the hydrogen bonding model used here, see a recent paper by Nguyen and Hall. For simplicity, the strength of a hydrophobic contact, $\epsilon_{HP}$, is fixed at 1/10 the strength of a hydrogen bond, $\epsilon_{HB}$. Hydrogen bond strength and hydrophobic contact strength are independent of temperature, as has been assumed in previous simulation studies.

6.2.2 Discontinuous Molecular Dynamics

Simulations are performed using the discontinuous molecular dynamics (DMD) simulation algorithm, which is an extremely fast alternative to traditional molecular dynamics and is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials. DMD simulations are conducted as follows. Each bead of the model protein chain is assigned a random initial position and a random initial velocity that do not violate any of the size constraints or assigned bond lengths and angles. The initial velocities are chosen at random from a Maxwell-Boltzmann distribution at a specified reduced temperature $T^* = k_B T / \epsilon_{HB}$, where $k_B$ is Boltzmann’s constant, $T$ is temperature, and $\epsilon_{HB}$ is the strength of the hydrogen bond in the model as explained earlier. When a DMD simulation begins, each bead moves with its individual velocity. The simulation proceeds according to the following schedule: identify the first event, move forward in time until that event occurs, calculate new velocities for the pair of beads involved in the event and calculate any changes in system energy resulting from hydrogen bond events or hydrophobic interactions, find the second event, and so on. Types of events include excluded volume events, bond events, and square-well hydrogen bond and hydrophobic interaction events. An excluded volume event occurs when the surfaces
of two hard-sphere beads collide and repel each other. A bond (or pseudobond) event occurs via a hard-sphere repulsion when two adjacent spheres attempt to move outside of their assigned bond length. Square-well events include well-capture, well-bounce, and well-dissociation “collisions” when a sphere enters, attempts to leave, or leaves the square well of another sphere. For more details on DMD simulations with square-well potentials, see papers by Alder and Wainwright\cite{44} and Smith, Hall, and Freeman\cite{61}.

Simulations are performed in the canonical ensemble which means that the number of particles, volume, and temperature are held constant. Periodic boundary conditions are used to eliminate artifacts due to simulation box wall effects. The dimensions of the box are chosen to ensure that a chain cannot interact with more than one image of any other chain. For this study, we use a cubic box with sides 100Å in length. Constant temperature is achieved by implementing the Andersen thermostat method\cite{62} as was used previously\cite{40,50}. With this procedure, all beads in the simulation are subject to random collisions with ghost particles. The post-event velocity of a bead colliding with a ghost particle is chosen randomly from a Maxwell-Boltzmann distribution at the simulation temperature.

6.2.3 Replica-exchange DMD/MC method

The replica-exchange method is implemented for four 96-peptide simulations conducted at concentrations $c = 0.5, 1, 2,$ and $5\text{mM}$, which range from the very dilute regime, in which most peptides do not interact with neighboring peptides, to the highly concentrated regime, in which most peptides are in contact with neighboring peptides. Each simulation at a concentration has 32 replica systems distributed over a broad interval of temperature ranging from $T^* = 0.10$ to
a high temperature at which each peptide is a random coil. Each replica system is simulated at a different temperature $T$ in the canonical ensemble using the DMD method. The number of replicas and the distribution of temperatures are chosen to ensure that: (a) there is a free random walk in temperature space, which means that every replica has the same probability of being switched to a neighboring temperature; (b) the number of replicas and hence temperatures sampled must be high enough to ensure that the probability of each replica being switched to a neighboring temperature is greater than 10%; and (c) the highest temperature sampled must be high enough to prevent the system from becoming trapped in a local energy minimum. These requirements are the same as those stated by Sugita and Okamoto $^{48}$. 

At fixed time intervals, replicas are sorted from lowest to highest temperature and subjected to the following temperature exchange MC procedure. Systems $i$ and $j$, with neighboring temperatures $T_i$ and $T_j$, respectively, can exchange configurations (system $i$ changes to temperature $T_i$ and system $j$ to temperature $T_j$) with probability:

$$
Probability = \begin{cases} 
1 & \text{if } \Delta \leq 0 \\
exp(-\Delta) & \text{if } \Delta > 0 
\end{cases}
$$

(6.1)

where $\Delta = [\beta_j - \beta_i](U_i - U_j)$ with $\beta_i = 1/(k_B T_i)$ and $U_i$ the potential energy of the system in state $i$. Initially each system is in a random configuration obtained from an NVT simulation at high temperature. Exchange attempts occur every $t^* = 0.5$ (at $c=0.5$mM), 0.4 (at $c=1.0$mM), 0.35 (at $c=2.0$mM), and 0.3 (at $c=5.0$mM) reduced time units. The reduced time $t^* \equiv t/\sqrt{k_B T/m}$ with $t$ the simulation time, and $\sigma$ and $m$ the average united atom diameter and mass. This corresponds to a replica-exchange attempt after approximately fifty million collisions at each
temperature. Approximately one thousand replica-exchange attempts are made during our simulations before equilibrium is reached. The criteria for equilibrium is that the ensemble average of the system’s total potential energy, which is collected at the end of each DMD run, should vary by no more than 2.5% during the second half of all DMD runs at each temperature.

Once equilibrium is reached, the data collection phase begins in which three hundred extra replica-exchange attempts are made. During the data collection phase, the properties of interest at each temperature are calculated throughout each DMD run. In addition, the probability of being in different energy levels and states is stored for conformational free energy calculations of the different thermodynamic properties of interest using the weighted histogram method. At the end of the replica-exchange DMD/MC simulation, our data contain a large ensemble of peptide configurations at each temperature and peptide concentration. Our simulations last more than sixty billion collisions at each temperature. A replica-exchange DMD/MC simulation at a single hydrophobic interaction strength and concentration requires 36 days on a cluster of 16 2.8-Ghz Xeon processors.

Our results are reported in part in terms of the average percentage of peptides in the system that form different structures which are defined in the following way. If 12 intra-peptide $\alpha$-helical hydrogen bonds (defined as bonds between $N_{i+4}$ and $C_i$) are formed, the structure is an $\alpha$-helix. If each peptide in a group of peptides has at least two inter-peptide hydrogen bonds or hydrophobic interactions with a neighboring peptide in the same group, then that group is classified as an aggregate. Aggregates can be either ordered or amorphous. If an aggregate contains $\beta$-sheets or better yet fibrils, we classify it as an ordered aggregate; otherwise, we classify it as an amorphous aggregate. If each peptide in a group of peptides has at least seven
inter-peptide $\beta$ hydrogen bonds to a particular neighboring peptide in the group, we classify this group as a $\beta$-sheet. If at least two $\beta$-sheet structures form inter-sheet hydrophobic interactions (at least four hydrophobic interaction per peptide per $\beta$-sheet) and the $\beta$-sheet structures are at an angle that is less than $35^\circ$, we classify this as a fibril; otherwise, we classify this and isolated $\beta$-sheets as unfibrillar $\beta$-sheet structures. A single pepti
de can be in a "$\beta$-structure" defined to be a structure that has a significant number of $\beta$ hydrogen bonds (a hydrogen bond between two residues whose backbone angles are in the $\beta$-region of the Ramachandran plot). Examples of single-peptide $\beta$-structures are $\beta$-hairpins and $\beta$-turns as defined in our previous paper\footnote{43}. Single-peptide structures that are not $\alpha$-helices or $\beta$-structures but have a small number of intra-peptide hydrogen bonds or hydrophobic interactions are random coils. The structures of particular interest in this study are fibrils, unfibrillar $\beta$-sheets, amorphous aggregates, $\alpha$-helices, $\beta$-hairpins, and random-coils.

To determine the locations of the thermodynamic transitions, the average radius of gyration $R_g$, the reduced specific heat $C_V^*$, and the potential energy $E$ are determined. The potential energy of the system $E$ is the sum of the energy contributed by hydrogen bonds (the number of hydrogen bonds times $\epsilon_{HB}$) and the energy contributed by hydrophobic interactions (the number of hydrophobic interactions times $\epsilon_{HP}$). The reduced specific heat, $C_V^*$, is calculated from the average potential energy $<E>$ and the average squared potential energy $<E^2>$:

\[
C_V^* = \frac{<E^2> - <E>^2}{k_B T^2}
\]  
(6.2)
6.3 Results and Discussion

Since this paper builds upon our previous works \(^{39,63}\) on the fibril formation of peptides of the same sequence, it is useful to briefly review those results that are pertinent to the discussion here. We investigated how peptide concentration and temperature affect the formation of various Ac-\(\text{KA}_{14}\)-K-NH\(_2\) structures including \(\alpha\)-helices, \(\beta\)-sheets and fibrils. Simulations were conducted on systems of 12, 24, 48, and 96 model 16-residue peptides at a wide variety of concentrations and temperatures by applying the discontinuous molecular dynamics simulation algorithm to our intermediate-resolution protein model. All simulations were performed in the canonical ensemble starting from a random coil configuration equilibrated at a high temperature and then slowly cooled to the temperature of interest so as to minimize kinetic trapping in local free energy minima. Structural characteristics such as the peptide arrangement and packing within fibrils were examined and compared with those observed in experiments. We also studied the overall stability of fibrils by conducting simulations on already-formed fibrils over a wide range of temperatures to investigate the relative importance of hydrogen bonding and hydrophobic interactions on fibril stability. The stability of our fibrillar structures was evaluated by comparing the abilities of the system to maintain the fibrillar structures at various temperatures which were higher than the fibril formation temperature.

We were able to observe the formation of fibrils containing 12 to 96 polyalanine peptides starting from random coils in a relatively short period of time ranging between 40 to 160 hours on a single processor of an AMD Athlon MP 2200+ workstation. To our knowledge, these were the first simulations to span the whole process of fibril formation from the random coil state to the fibril state on such a large system. We found that there was a strong relationship between the
formation of $\alpha$-helices, $\beta$-sheets, aggregates and fibrils and the environmental conditions such as temperature, concentration, and hydrophobic interaction strength. At low concentrations, $c < 1.0$ mM, random-coil peptides assembled into $\alpha$-helices at low temperatures and random coils at high temperatures, $T^* > 0.10$. At intermediate concentrations, $c = 1.0 - 2.5$ mM, random-coil peptides assembled into $\alpha$-helices at low temperatures, $T^* = 0.08 - 0.10$, and large $\beta$-sheet structures at high temperatures, $T^* = 0.11 - 0.14$. At high concentrations, $c > 2.5$ mM, random-coil peptides formed $\beta$-sheets over a wide range of temperatures, $T^* = 0.08 - 0.14$. These $\beta$-sheets assembled into fibrils above a critical temperature that decreased with concentration and exceeded the isolated peptide’s folding temperature ($T^* = 0.11$). At very high temperatures, $T^* \geq 0.15$, and all concentrations, the system was in a random-coil state. The critical concentration for fibril formation was between $c = 1.0$ mM and $c = 2.5$ mM at high temperatures $T^* = 0.12 - 0.14$. These results were in good qualitative agreement with the experimental results of Blondelle et al. on Ac-KA$_{14}$K-NH$_2$ peptides$^{52,53}$. The fibrils observed in our simulations mimicked the structural characteristics observed in experiments in that most of the peptides in our fibrils were arranged in an in-register parallel orientation, with intra-sheet and inter-sheet distances similar to those observed in experiments, and contained about six multi-peptide $\beta$-sheets. Finally, we found that when the strength of the hydrophobic interaction between non-polar sidechains relative to the strength of hydrogen bonding was increased from $R = 1/10$ to $R = 1/6$, the system formed amorphous rather than fibrillar aggregates. We also identified key fibril-forming events. Since simulations were conducted by cooling the system down to the temperature of interest, analysis of the temperature-dependence of the kinetics of fibril formation was not appropriate.

We then investigated the kinetics of fibril formation of Ac-KA$_{14}$K-NH$_2$ peptides as a func-
tion of the peptide concentration and temperature. Simulations were conducted on systems containing 48 model 16-residue peptides in the canonical ensemble at a wide variety of concentrations and temperatures. During each simulation, the formation of different structures such as α-helices, amorphous aggregates, β-sheets or fibrils was monitored as a function of time. Key fibril-forming events that were associated with the four proposed fibril forming mechanisms mentioned earlier are identified. Two types of simulations were conducted: unseeded and seeded. In the unseeded simulations, the peptides, which were initially in a random coil configuration, were equilibrated at a high temperature and then quickly cooled to the temperature of interest. The lag time and rate of fibril formation were monitored to study their dependence upon the peptide concentration and temperature. In the seeded simulations, a previously created single fibril was immersed in a sea of denatured chains. The single-chain docking process on the molecular level was monitored, the conformation of the chain before or after it encounters the fibril surface was identified, and the size distribution of the different structures as a function of time was determined. Our results show that fibril formation for polyalanines incorporates different features that are characteristic of three models: templated assembly, nucleated polymerization, and nucleated conformational conversion. Fibril formation is nucleation dependent with an ordered nucleus of two β-sheets, each with two to three peptides. The lag time before fibril formation commences decreases with increasing concentration and increases with increasing temperature. In addition, fibril formation appears to be a nucleated conformational conversion process in which small amorphous aggregates → β-sheets → ordered nucleus → subsequent rapid growth of a stable fibril. Fibril growth in our simulations involves both β-sheet elongation, in which the fibril grows by adding individual peptides to the end of each β-sheet and lateral addition, in which the
fibril grows by adding already-formed β-sheets to its side. Moreover, the rate of fibril formation increases with increasing concentration and decreases with increasing temperature.

In this section, we present the results from equilibrium (replica-exchange) simulations of 96-peptide systems. At low concentrations, the system goes from a one-phase region containing α-helices to a two-phase region containing both unfibrillar β-sheets and random coils and then to a one-phase region containing random coils as can be seen in Figure 6.2, which plots the percentage of peptides in different structures as a function of the reduced temperature $T^*$ for the 96-peptide system at $c=0.5$mM. This figure indicates that at low temperatures ($T^* = 0.09 - 0.10$), the vast majority of peptides form α-helices compared to that of β-hairpins as expected based on the intrinsic α-helical property of polyalanines in dilute solutions. The temperature at which half of the peptides form α-helices is $T^* = 0.110$, which is the midpoint of the folding transition (50% helicity) of a single peptide from our previous simulations. As the temperature increase to intermediate temperatures ($T^* = 0.110 - 0.135$), the system goes to a two-phase region that has predominantly random-coils and less prominently unfibrillar β-sheets. The formation of β-sheets at intermediate temperatures is also observed for single peptides based on our previous simulations. At high temperatures ($T^* > 0.135$), the only structure that appears is the random coil.

The transition between a one-phase region containing α-helices to a two-phase region containing both unfibrillar β-sheets and random coils can be seen in Figure 6.3, which plots the reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) as a function of the reduced temperature $T^*$ for the same system as in Figure 6.2. The transition temperature can be identified from the peaks in the specific heat $C_V^*$, which is the slope of the potential energy with respect to the
temperature. Although the thermodynamic properties $C_V^*$ and $R_g$ show the transition between an one-phase region to a two-phase region at $T^* = 0.110$, they do not show the transition at a high temperature between the two-phase region containing unfibrillar $\beta$-sheet and random coil to the one-phase region containing random coils. In Figure 6.3, the transition that can be seen is at $T^* = 0.110$, which is the same as the midpoint of the $\alpha$-helical folding transition at $T^* = 0.11$ deduced from Figure 6.2. The radius of gyration also reflects the phase transition. For example, the radius of gyration in the one-phase region which contains $\alpha$-helices ($T^* = 0.09$) is 7.31Å, which is comparable to 7.27Å for a perfect $\alpha$-helix. In the two-phase region at $T^* = 0.125$ which contains both random coils and unfibrillar $\beta$-sheets, the radius of gyration is 11.77Å, which is between the value of 10.45Å for a single random-coil conformation and 13.10Å for an extended peptide conformation such as those observed in $\beta$-sheets from our previous simulation studies. In the one-phase random coil region at $T^* > 0.145$, the radius of gyration is 11.41Å, which is higher than the value for a typical random coil.

As the concentration increases from $c=0.5$mM to $c=1.0$mM, the transition between different phases is hard to detect since at each temperature, the system contains more than one structural state as can be seen in Figure 6.4, which plots the percentage of peptides in different structures as a function of the reduced temperature $T^*$ for the 96-peptide system at $c=1.0$mM. At low temperatures ($T^* < 0.095$), the structural state that has the highest number of peptides is the $\alpha$-helical structure at about 40%, followed by the unfibrillar $\beta$-sheet at about 30% and the amorphous aggregate at about 25%. As the temperature increases from $T^* = 0.095$, the percentages of peptides that form $\alpha$-helices and amorphous aggregates decrease. In contrast, the percentage of peptides that form unfibrillar $\beta$-sheets increases to a maximum. Over the temperature range $T^* = 0.115$
- 0.125, the percentage of peptides that form unfibrillar $\beta$-sheets is relatively high with the peak of 60% at $T^* = 0.115$, at which the percentage of peptides that form each of the other structures such as fibrils, amorphous aggregates, and $\beta$-hairpins is about 10%. As the temperature increases from $T^* = 0.125$ to $T^* = 0.14$, the percentage of peptides that form fibrils increases, peaking at a value of 20%; over this temperature range only 5% of peptides form amorphous aggregates and the remaining 75% the peptides form random coils.

The thermodynamic properties $C_V^*$ and $R_g$ of the system at $c = 1.0\text{mM}$ are shown in Figure 6.5, which plots the reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) as a function of the reduced temperature $T^*$ for the same system as in Figure 6.4. The specific heat results show a peak at approximately $T^* = 0.128$ corresponds the midpoint of the unfibrillar $\beta$-sheet curve reflecting phase transition between a multiple-phase region at unfibrillar $\beta$-sheet is the dominant structure to a region that has predominantly random-coils and less prominently fibrils. The radius of gyration results show more phase transitions than the specific heat data. At $T^* < 0.095$, the radius of gyration is $R_g = 8.5$ Å which is leaning more to the value of a perfect $\alpha$-helix (7.27 Å) than to a somewhat random coil found in amorphous aggregates (10.45 Å) or a $\beta$-sheet conformation (13.10 Å). At $T^* = 0.10 - 0.11$, the radius of gyration increases to $R_g = 11.0$ Å, which is comparable to the value of a conformation in amorphous aggregates (10.45 Å) but actually is contributed by $\alpha$-helices and unfibrillar $\beta$-sheets since they are the most dominant structures in this multi-phase region. The change in the radius of gyration reflects the phase transition between multi-phase regions that are dominated by $\alpha$-helices at $T^* < 0.095$ and by unfibrillar $\beta$-sheets at $T^* = 0.10 - 0.11$. The radius of gyration then increases to $R_g = 12.5$ Å at $T^* = 0.125$ for the region at which a vast majority of peptides are unfibrillar $\beta$-sheets. The radius
of gyration then decreases to 11.0 Å at $T^* = 0.14$ for the random coil.

As the concentration increases from $c=1.0\, \text{mM}$ to $c=2.0\, \text{mM}$, there is only one transition between the fibrils and random coils as can be seen in Figure 6.6, which plots the percentage of peptides in different structures as a function of the reduced temperature $T^*$ for the 96-peptide system at $c=2.0\, \text{mM}$. At $T^* < 0.11$, most peptides are in fibrils and less peptides are in un fibrillar β-sheets. At $T^* = 0.11 \sim 0.13$, the population of fibrils is high. At $T^* > 0.13$, the percentage of peptides that form fibrils decreases as the percentage of peptides that form random coils increases.

The thermodynamic data shows the same phase transition as that from the data on the percentage of peptides that form the different structures as seen in Figure 6.7, which plots the reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) as a function of the reduced temperature $T^*$ for the same system as in Figure 6.6. The specific heat results show a peak at $T^* = 0.13$, which corresponds to the higher limit of the temperature region at which fibrils are at a maximum. The radius of gyration also a phase transition between between fibrils and random coils as its value drops from 13.2 Å at $T^* = 0.12$ to 11.0 Å at $T^* \geq 0.14$. In addition, the radius of gyration also shows an increase in the percentage of peptides that form fibrils over the lower range of the transition temperature as its value increases from 12.0 Å at $T^* = 0.09$ to 13.2 Å at $T^* = 0.13$.

As the concentration increases from $c=2.0\, \text{mM}$ to $c=5.0\, \text{mM}$, there is only one phase transition between the fibril and the random coil as can be seen in Figure 6.8, which plots the percentage of peptides in different structures as a function of the reduced temperature $T^*$ for the 96-peptide system at $c=5.0\, \text{mM}$. Over a wide temperature range from $T^* = 0.09$ to $T^* = 0.14$,
a high percentage (80%) of the peptides form fibrils and a low percentage (about 15%) form unfibrillar β-sheets. At $T^* > 0.14$, most peptides form random coils.

The thermodynamic properties $C_V^*$ and $R_g$ of the system at $c=5.0\text{mM}$ show the transition between the fibril and the random coil as seen in Figure 6.9, which plots the reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) as a function of the reduced temperature $T^*$ for the same system as in Figure 6.8. The transition is at approximately $T^* = 0.135$, which is slightly lower than the midpoint of the fibril transition at $T^* = 0.139$ as reflected in Figure 6.8.

The results for the 96-peptide system that we have just described are summarized in Figure 6.10, which shows the phases that occur in the space spanned by the reduced temperature $T^*$ and peptide concentration $c$. Here we call a particular structure a phase if the percentage of peptides that form that structure is significant (i.e. at least 50%). If the structure with the second highest percentage of the peptides has a percentage at least 20%, we then say that there are two phases. If no structure has a percentage of 50% or higher, then we consider the next two structures that have the highest percentages with the highest one considered as the most prominent structure in that phase. In other words, we consider phases that contain between one and two structures with percentages greater than or equal to 20%. Figure 6.10 indicates that there four single-phase regions: α-helices, fibrils, unfibrillar β-sheets and random coils. In addition, there are four different two-phase regions: random coils / unfibrillar β-sheets, random coils / fibrils, fibrils / unfibrillar β-sheets, and α-helices / unfibrillar β-sheets. Interestingly enough, a two-phase region containing fibrils and unfibrillar β-sheets is not observed indicating that most β-sheets that are formed in our simulations are assembled into fibrils.

Alpha-helices are formed at high percentages only at low temperatures ($T^* \leq 0.11$) and low
concentrations \((c \leq 0.5\text{mM})\). As the concentration is increased from \(c = 0.5\text{mM}\) to \(c = 1.0\text{mM}\), \(\alpha\)-helices are formed in lower percentages and at lower temperatures \((T^* \leq 0.10)\) than in the less concentrated case. At intermediate temperatures \((T^* = 0.12-0.13)\), unfibrillar \(\beta\)-sheets are formed at low concentrations \((c \leq 0.5\text{mM})\). As concentration is increased from \(c = 0.5\text{mM}\) to \(c = 1.0\text{mM}\), unfibrillar \(\beta\)-sheets are formed in higher percentages and at lower temperatures \((T^* = 0.09-0.13)\) than in the less concentrated case. As concentration is increased further from \(c = 1.0\text{mM}\) to \(c = 2.0\text{mM}\), fibrils are formed in high percentages at intermediate concentrations \((T^* = 0.11-0.13)\). As the concentration is increased from \(c = 2.0\text{mM}\), fibrils are formed in higher percentages and at lower and higher temperatures \((T^* = 0.09-0.14)\) than in the less concentrated case. At all concentrations and high temperatures \((T^* \geq 0.13)\), the system forms random coils.

6.4 Conclusions

In this paper, we performed equilibrium simulations on 96-peptide systems over a very wide range of temperatures and peptide concentrations by using the replica exchange simulation method. Based on the thermodynamic properties \(C_V^*\) and \(R_g\) of the system at each concentration and the data on the percentage of peptides that form the various structures, we mapped out a phase diagram in the temperature-concentration plane delineating the regions where different structures are stable. We found that there are four distinctive single-phase regions: \(\alpha\)-helices, fibrils, unfibrillar \(\beta\)-sheets and random coils. The \(\alpha\)-helical region occurs at low temperature and low concentration. The \(\beta\)-sheet structures that are not in fibrils are at intermediate temperatures; this \(\beta\)-sheet region expands to lower temperatures as concentration is increased. The fibril region
occurs mostly at intermediate temperatures and intermediate concentrations and expands to lower
and higher temperatures as the peptide concentration is increased. The random-coil region occurs
at high temperatures at all concentration and shifts to higher temperatures as the concentration
is increased. In addition, there are four different two-phase regions: random coils / unfibrillar
\( \beta \)-sheets, random coils / fibrils, fibrils / unfibrillar \( \beta \)-sheets, and \( \alpha \)-helices / unfibrillar \( \beta \)-sheets.

6.5 Acknowledgments

The authors are grateful to Sylvie Blondelle, Ken Dill, Jeffery Kelly, and Dan Kirschner for
helpful discussions. This work was supported by the National Institutes of Health under grant
number GM-56766 and the National Science Foundation under grant number CTS-9704044.
6.6 References


6.7 Figures

6.1 Geometry of the intermediate resolution protein model for polyalanine. Covalent bonds are shown with narrow solid lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive C\(_\alpha\) distances, and residue L-isomerization. Note that the united are not shown full size for ease of viewing.

6.2 The percentage of peptides in \(\alpha\)-helices (filled circles), fibrils (open squares), unfibrillar \(\beta\)-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=0.5\) mM.

6.3 Reduced specific heat \(C_V^*/T\) and radius of gyration \(R_g\) (in Å) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=0.5\) mM.

6.4 The percentage of peptides in \(\alpha\)-helices (filled circles), fibrils (open squares), unfibrillar \(\beta\)-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=1.0\) mM.

6.5 Reduced specific heat \(C_V^*/T\) and radius of gyration \(R_g\) (in Å) versus the reduced temperature \(T^*\) for the 96-peptide system at \(c=1.0\) mM.
6.6 The percentage of peptides in α-helices (filled circles), fibrils (open squares),
unfibrillar β-sheets (filled diamonds), amorphous aggregates (opened triangles),
hairpins (filled triangles), and random coils (stars) versus the reduced tempera-
ture $T^*$ for the 96-peptide system at $c=2.0\text{mM}$. . . . . . . . . . . . . . . . . . 271

6.7 Reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) versus the reduced
temperature $T^*$ for the 96-peptide system at $c=2.0\text{mM}$. . . . . . . . . . . . . . 272

6.8 The percentage of peptides in α-helices (filled circles), fibrils (open squares),
unfibrillar β-sheets (filled diamonds), amorphous aggregates (opened triangles),
hairpins (filled triangles), and random coils (stars) versus the reduced tempera-
ture $T^*$ for the 96-peptide system at $c=5.0\text{mM}$. . . . . . . . . . . . . . . . . . 273

6.9 Reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) versus the reduced
temperature $T^*$ for the 96-peptide system at $c=5.0\text{mM}$. . . . . . . . . . . . . . 274

6.10 Phase diagram for the 96-peptide system as a function of the reduced temper-
ature, $T^*$, and peptide concentration, $c$. The one-structure phases are α-helices
(black circles), fibrils (red circles), unfibrillar β-sheets (blue circles), and random
coils (purple circles). The two-structure phases are represented by squares with
the structure that has the highest percentage is listed first; they are random coils /
unfibrillar β-sheets, random coils / fibrils, fibrils / unfibrillar β-sheets, α-helices
/ unfibrillar β-sheets, and unfibrillar β-sheets / α-helices. . . . . . . . . . . . . . . 275
Figure 6.1: Geometry of the intermediate resolution protein model for polyalanine. Covalent bonds are shown with narrow solid lines connecting beads. At least one of each type of pseudobond is shown with a thick disjointed line. Pseudobonds are used to maintain backbone bond angles, consecutive $C_\alpha$ distances, and residue L-isomerization. Note that the united are not shown full size for ease of viewing.
Figure 6.2: The percentage of peptides in α-helices (filled circles), fibrils (open squares), unfibrillar β-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature $T^*$ for the 96-peptide system at $c=0.5$mM.
Figure 6.3: Reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) versus the reduced temperature $T^*$ for the 96-peptide system at $c=0.5$ mM.
Figure 6.4: The percentage of peptides in α-helices (filled circles), fibrils (open squares), unfibrillar β-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature $T^*$ for the 96-peptide system at $c=1.0\text{mM}$. 
Figure 6.5: Reduced specific heat $C^*_V$ and radius of gyration $R_g$ (in Å) versus the reduced temperature $T^*$ for the 96-peptide system at $c=1.0$ mM.
Figure 6.6: The percentage of peptides in $\alpha$-helices (filled circles), fibrils (open squares), unfibrillar $\beta$-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature $T^*$ for the 96-peptide system at $c=2.0\text{mM}$. 
Figure 6.7: Reduced specific heat $C^*_V$ and radius of gyration $R_g$ (in Å) versus the reduced temperature $T^*$ for the 96-peptide system at $c=2.0$ mM.
Figure 6.8: The percentage of peptides in α-helices (filled circles), fibrils (open squares), unfibrillar β-sheets (filled diamonds), amorphous aggregates (opened triangles), hairpins (filled triangles), and random coils (stars) versus the reduced temperature $T^*$ for the 96-peptide system at $c=5.0\text{mM}$. 
Figure 6.9: Reduced specific heat $C_V^*$ and radius of gyration $R_g$ (in Å) versus the reduced temperature $T^*$ for the 96-peptide system at $c=5.0$ mM.
Figure 6.10: Phase diagram for the 96-peptide system as a function of the reduced temperature, $T^*$, and peptide concentration, $c$. The one-structure phases are $\alpha$-helices (black circles), fibrils (red circles), unfibrillar $\beta$-sheets (blue circles), and random coils (purple circles). The two-structure phases are represented by squares with the structure that has the highest percentage is listed first; they are random coils / unfibrillar $\beta$-sheets, random coils / fibrils, fibrils / unfibrillar $\beta$-sheets, $\alpha$-helices / unfibrillar $\beta$-sheets, and unfibrillar $\beta$-sheets / $\alpha$-helices.
CHAPTER 7

FUTURE WORK

The work in this thesis on fibril formation by polyalanines shows the power and robust nature of the discontinuous molecular dynamics simulation algorithm, along with the simplicity of our intermediate-resolution protein model. When combined, they enable the observation of the complete fibrillization process from random initial configurations in very large systems within reasonable times. In contrast, many investigators have applied atomic resolution computer simulation to systems containing either a few small proteins or an already-formed fibril but computer limitations prevent them from observing fibril formation. This suggests that our current model and method should be extended to the study of fibril formation by many peptides other than the alanine-based sequences. At the same time, there are several areas in which the work on alanine-based peptides should be continued.

7.1 Fibril formation by alanine-based peptides

Now that the solution conditions such as the peptide concentration and temperature for fibril formation at which Ac-KA$_{14}$K-NH$_2$ peptides occurs have been determined, we could learn more about the mechanisms that would prevent fibril formation. We could begin by simulating the
polyalanine fibrillation process in the presence of model inhibitors, which could be modeled as very short polyalanine peptides. This would be in keeping with the approach now being taken in the \(\beta\)-amyloid community where short hydrophobic stretches of the peptide in question are used as inhibitors \(^1\).

We could also study the effects of varying the peptide chain length on the formation of fibrils by alanine-based peptides and compare our results with the experimental data obtained by Blondelle \textit{et al.}\(^2\), who determined the critical chain length for fibril formation by peptides of the Ac-KA\(_n\)K-NH\(_2\). They found that peptides shorter than \(n=9\), where \(n\) is the number of alanine residue peptides, did not form \(\beta\)-sheet complexes, those between \(n=9\) and \(n=14\) formed both \(\alpha\)-helices and \(\beta\)-sheet complexes, and those longer than \(n=14\) formed \(\beta\)-sheet complexes.

We could also examine the effects of varying the ratio of the number of hydrophobic residues relative to the number of hydrophilic residues and the position of the hydrophobic residues along the chain. We could compare our results against experimental data on short sequences that have been found to form fibrils. One example would be the designed combinatorial libraries of Hecht \textit{et al.}\(^3\), who show that alternating polar and nonpolar residues along the sequences predisposes a sequence to form amyloid-like structures. We could also compare our results against experimental data obtained by Wang \textit{et al.}\(^4\), who also study the role of hydrophobic residues in the assembly of different artificial amphiphilic peptides that are derived from the FEFEFKFK sequence through substitution experiments. They find that fibril formation is independent of the type of amino acid that is used for substitution as long as the hydrophobic residues are at least in alternating positions along the peptide. The only difference from sequence to sequence that they find is in the morphology of the resultant self-assembled ma-
trix; this is caused by the sizes of the hydrophobic sidechains. By this logic, we could use the current protein model and vary the number, position, and size of hydrophobic sidechains on the Ac-KA_{14}K-NH_2 peptide to investigate the effects of varying different properties of hydrophobic residues on the morphological difference in the resultant fibrils.

We could investigate fibril formation by AGAAAAGA, a small segment of the Syrian hamster prion protein, which is associated with prion diseases. Modeling AGAAAAGA is feasible since our current protein model can handle glycines. It would be interesting to repeat all of simulations that we have conducted on polyalanines and compare the results with those from experiments.

7.2 Protein Model

Current efforts in our group are underway to extend the current intermediate-resolution protein model that has already been developed for polyalanine to the description of other homopeptides. This is being done in a series of steps. The model is being modified to represent the geometry and energetics of polyvaline first, then polyleucine, polyisoleucine, and finally polyglutamine whose extended tracts have been implicated in Huntington’s disease. In extending our current intermediate-resolution protein model to the description of other amino acids, we need to modify the sidechain representation since the backbone is already very detailed. Once this approach has been finished, we could extend the new protein model to the description of all twenty amino acids so that it can be applied to other heteropeptides that are associated with various neurodegenerative diseases. We need to choose a side-chain representation and interaction potential model that are realistic enough to distinguish the role played by each type of residue in
aggregate formation yet simple enough to allow the simulation of fibril formation in a reasonable time frame. We also need to choose a method for testing our predictions against experimental data.

There are many different ways to extend our intermediate resolution protein model to the description of all 20 amino acids. Although we could take a fairly detailed approach and create an appropriate multi-bead representation for each of the 20 side chains, this would slow the code down to the point where we would lose our ability to access fibrillization time scales. Instead, our approach should be to make the model for each side chain as simple as we can, without neglecting the unique role played by each side chain in allowing or inhibiting fibrillization. Accordingly, the first model that we should consider would be a single-bead side chain model in which the side chain size, distance from C\text{\textalpha}, and hydrophobicity of each residue are unique. Prolines could be handled by adding a pseudobond linking the single-bead side chain to the backbone NH. Charged residues would be represented by a (possibly multi-step) square-well or square-shoulder potential of long range to mimic a screened Coulomb potential between like and unlike charges and account for salt bridges.

### 7.3 Fibril formation by heteropeptides

The new 20-amino-acid model could be tested by comparing its fibrillization predictions to results appearing in the literature on the fibril-forming propensity of short peptides. We could start by repeating the substitution experiments of Blondelle et al. on Ac-KA\textsubscript{n}K-NH\textsubscript{2} to see if our fibrillization predictions are in agreement with hers. By substituting proline at various positions along a 14-mer chain, they showed that complex formation was not inhibited when the proline
was substituted in the region alanine-7 to alanine-11. Substitution of the other nineteen amino acid residues at alanine-10 showed that side chains bulkier than alanine (methionine, leucine or isoleucine) and charged amino acid residues disrupted β-sheet formation; glycine, which is thought to destabilize secondary structure, did not. Blondelle et al. also showed that the tendencies of each amino acid to form β-sheets did not correlate with its β-sheet propensity.

We could also compare our 20-amino-acid model’s fibrillization predictions to results obtained by many studies in the literature that describe which fragments of disease-related amyloidogenic proteins form fibrils and which do not. Examples of very short amyloidogenic protein fragments that fibrillize and their parent protein are the following: FGAIL and NF-GAIL (amylin) 7 and VQIVYK (tau protein) 8. Examples of peptides that do not fibrillize are NGKSNFLNCYVSG and FHPSDIEVDLLK, which are derived from from the protein B₂-microglobulin 9. Once the new 20-amino-acid model is tested on short peptides, we could start investigating fibril formation by longer peptides that are implicated in many diseases.

We could also simulate the assembly of non-disease related peptides into ordered structures with the ultimate goal of observing the formation of biomaterials, which have received considerable attention lately due to their outstanding mechanical, electronic, magnetic or optical properties 10–13. For example, biomolecular materials based upon natural silks are light and soft materials yet they are stronger than steel and Kevlar 14. They derive their extraordinary strength from hard, crystalline antiparallel β-sheet structures which contain a highly-dense network of hydrogen bonds and their elasticity from amorphous peptide regions which contain bulky side groups and unordered conformations 14. In addition, biomaterials produced by short sequences of natural amino acids are bio-degradable, unlike other synthetic materials that are
produced by such polymers as polyethylene and poly(lactic acid)\(^\text{15}\). Moreover, certain properties of oligopeptide biomaterials can be tailored to specific applications. For example, increasing the concentration of the oligopeptides FKFEFKFE can change the mechanical properties of the resultant gel, specifically increasing the Young’s modulus of the final structure\(^\text{16}\). This oligopeptide FKFEFKFE could be easily modeled with our 20-amino-acid model.

We could also model other non-disease related short peptides that are known to form ordered aggregates: (1) the silk-elastin polymers, which are block copolymers containing a silk-like block GAGAGS that spontaneously self assembles and an elastin-like block GVGVP that inhibits self assembly\(^\text{12}\), (2) the peptides P11-I (QQRQQQQQQQQ) and P11-II (QQRFWQFQQQ) of Aggeli et al. which form tapes ribbons and fibrils\(^\text{17}\), (3) the self-assembling peptides of Zhang et al. which contain combinations of the repeats RADA, RARA, DADA, AEAK, RAHA, KADA and many others\(^\text{18}\), and (4) peptide fragments from the wild-type hen lysozyme (GST-DYGILQINSRWWS)\(^\text{19}\). We could also model spider dragline silk proteins, Spidroin I and II, which are block copolymers of eight-alanine blocks and of glycine-rich domains\(^\text{20}\). Finally, we could simulate the self-assembly of β-sheets into nanostructures by block copolymers containing alternating segments of either poly(A) or poly(AG) and segments of poly(ethylene glycol)\(^\text{21,22}\), which produces nanostructures that are structurally similar to spider silk\(^\text{14}\) or \textit{B. mori} silk\(^\text{22}\).

In simulating the aggregation of different peptides, we could study the effects of the peptide sequence, concentration, solvent condition, and temperature on the self-assembly process. In the disease-related case, we could determine the factors important in fibril formation so as to formulate mechanisms that could prevent fibril formation. In the non-disease-related case, we could determine the factors important in forming biostructures with superior properties so as to
formulate mechanisms that could help fibril formation. This could lead to some valuable insights into new therapeutic strategy for amyloidogenic diseases as well as the design and synthesis of biomaterials with controlled architecture.
7.4 References


