Structure and Thermodynamics of Polyglutamine Peptides and Amyloid Fibrils via Metadynamics and Molecular Dynamics Simulations

Riley Workman

Follow this and additional works at: https://dsc.duq.edu/etd

Part of the Biological and Chemical Physics Commons, Biophysics Commons, and the Physical Chemistry Commons

Recommended Citation


This One-year Embargo is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection. For more information, please contact phillipsg@duq.edu.
STRUCTURE AND THERMODYNAMICS OF POLYGLUTAMINE PEPTIDES AND AMYLOID FIBRILS VIA METADYNAMICS AND MOLECULAR DYNAMICS SIMULATIONS

A Dissertation
Submitted to the Bayer School of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for
the degree of Doctor of Philosophy

By
Riley J. Workman

August 2018
STRUCTURE AND THERMODYNAMICS OF POLYGLUTAMINE PEPTIDES
AND AMYLOID FIBRILS VIA METADYNAMICS AND MOLECULAR
DYNAMICS SIMULATIONS

By
Riley J. Workman

Approved July 6th, 2018

Jeffrey D. Evanseck
Professor of Chemistry
(Committee Chair)

Mihaela Rita Mihailescu
Professor
(Committee Member)

Phillip J. Reeder
Dean, Bayer School of Natural and Environmental Sciences
Dean and Professor

Ellen S. Gawalt
Professor
(Committee Member)

Matthew N. Srnec
Assistant Professor
Franciscan University of Steubenville
(External Committee Member)

Ellen S. Gawalt
Chair, Department of Chemistry and Biochemistry
Professor
ABSTRACT

STRUCTURE AND THERMODYNAMICS OF POLYGLUTAMINE PEPTIDES
AND AMYLOID FIBRILS VIA METADYNAMICS AND MOLECULAR
DYNAMICS SIMULATIONS

By
Riley J. Workman
August 2018

Dissertation supervised by Jeffrey D. Evanseck

Aggregation of polyglutamine (polyQ)-rich polypeptides in neurons is a marker for nine neurodegenerative diseases. The molecular process responsible for the formation of polyQ fibrils is not well understood and represents a growing area of study. To enable development of treatments that could interfere with aggregation of polyQ peptides, it is crucial to understand the molecular mechanisms by which polyQ peptides aggregate into fibrils. Many experimental techniques have been employed to probe polyQ aggregation, however, observations from these studies have not lead to a unified understanding of the properties of these systems, instead yielding competing, fragmented theories of polyQ aggregation. This dissertation addresses these gaps in knowledge by shedding light on important steps of the aggregation process. The structural motif of polyQ fibrils is not agreed upon in the field, which is worrying, given that these structures are the endpoint of polyQ aggregation. Here, molecular dynamics (MD) simulations paired with UV resonance Raman (UVRR) experiments
show that short polyQ peptides adopt extended antiparallel β-sheet fibrils, contrary to β-hairpin structures oft predicted in the polyQ field. The structure of monomeric polyQ peptides was then studied to gain insight into the beginnings of the aggregation mechanism. Metadynamics MD simulations were used to characterize the conformational energy landscape of polyQ peptides, and this data was compared to experimental UVRR results. We found short polyQ peptides can adopt PPII-rich and collapsed β-strand monomeric structures, which establishes that polyQ can form distinct conformational states as monomers. The effect of increased polyQ repeat length was also tested, and it was found that increased repeat length corresponds to lower energy barriers between monomeric conformational states, which may explain why longer polyQ repeats are quicker to aggregate. Hydrogen bonding strengths of polyQ monomers and fibrils were also investigated with MD and UVRR, showing that polyQ peptides favor intrapeptide hydrogen bonds over those between peptide and water. Overall, the work in this dissertation deepens the understanding of the polyQ aggregation mechanism by determining the structure and thermodynamics of monomeric and fibrillar states, as well as identifying polyQ peptide hydrogen bonding as one of the driving forces in these systems. This knowledge can aid the development of molecular mechanisms to interfere with the formation of toxic polyQ aggregates that trigger the onset of polyQ diseases.
DEDICATION

This dissertation is dedicated to my late research advisor, mentor, and friend, Dr. Jeffry D. Madura. Jeff was a gentle, kind, and fair man who always found the good in everyone, even if I couldn’t see it. He had genuine, infectious passion for the things he involved himself with, which included teaching science, doing science, eating good food, watching hockey, and traveling. When I joined Jeff’s lab I might’ve only claimed to enjoy one or two of those things, but now I’d say, more or less, I enjoy all of them. Jeff introduced me to many of life’s great pleasures, such as Bahamian rum, cheesecake, fresh guacamole, Python programming, and partial molar volume (whatever that means). More importantly, he was the only person that could keep up with my walking pace, which came in handy searching for the best places to eat in San Francisco, New Orleans, Los Angeles, San Diego, and Baltimore. That was just the kind of guy that he was, his fascination, curiosity, and joy were addicting. Jeff taught me to never shy away from a new scientific direction or project, and, to my possible detriment, I have heeded that advice. We should all aim to take every day the way Dr. Madura did: as a chance to teach, love, learn, and laugh. Jeff was taken from us far too soon, but the countless students, coworkers, friends, and loved ones he inspired will carry on his legacy. I know he would be proud to see me graduate, and I am proud to have been his student and friend. Jeff, thanks for everything. I miss you every day.
ACKNOWLEDGEMENT

I’d like to thank all of the people who helped me throughout my six years at Duquesne University. Without the help, love, care, support, friendship, and general all around life talent of my friends from the Madura and Evanseck research groups such as Emily Benner, Kendy Pellegrene, Matt Srnec, Shiv Upadhyay, Amanda Dumi, Mike Wasko, Gabby Pros, and Amy Carlson, I surely would not have made it through this ordeal. Emily was my consummate second in all things at Duquesne and elsewhere during my graduate career. She understands me as the weird person I am, and that is enough. Kendy calmed me down when I was anxious and made me laugh, even though I am funnier than her. For my first four years at Duquesne, Matt was basically my work dad, and in a turn of fate, also became my external reader. Shiv has been a great friend, roommate, and CS:GO teammate, and Amanda is basically the only person who laughs at my jokes. Mike Wasko is essentially the nicest person in the world... I’m not really even exaggerating, I love that guy. Gabby and Amy are friends whose conversation, laughter, and company I value dearly.

I would be remiss not to mention the great faculty members that helped me along the way. Firstly, I would like to thank Dr. Jeffrey Evanseck for being a great boss. When things were difficult, he took me under his wing and gave me the guidance I needed to graduate. Dr. Rita Mihailescu was a great mom away from home when I was at various Biophysical Society meetings, and she is also a scientist and person that I greatly admire. Dr. Michael Cascio is great fun, and always led me to the primo bars and restaurants in conference towns. I would like to thank Dr. Ellen Gawalt for stepping up when I needed her and being a great committee member. My collaborator at the University of Pittsburgh, Dr. Sandy Asher, was basically a co-PI in most of my projects, and I wouldn’t have been as successful without him. I learned a lot about being a hip college professor from Dr. Mary Sherman, and she was always
very helpful. These are not the only faculty members I need to thank; there are many other professors who helped me along the way, including but not limited to: Dr. Philip Auron, Dr. Ralph Wheeler, Dr. Paul Johnson, Mitch Fedak, Dr. Nithya Vaidyanathan, Dr. Doug Fox, and Dr. Emilio Esposito.

There are several other staff members at Duquesne who deserve a shoutout. Firstly, not enough can be said about Scott Boesch. He has always been a kind, caring, and fun influence at work, and went out of his way to help me on many occasions. My time at Duquesne was enriched by his presence and friendship. I’d like to thank Amy Stroyne, Maggie Cowburn, Aimee Ciesielski, Lisa Zandier, and Heather Costello, who have all helped keep my life organized(ish) throughout my graduate career.

Throughout my time in graduate school I collaborated with some excellent young scientists from the University of Pittsburgh, namely Dr. Dave Punihaoale, Ryan Jakubek, and Steve White. Dave was the first student of Dr. Asher’s that I collaborated with, and he has been a great friend and research partner for the last five years. I have great respect for Dave as a guy and a scientist, and I will hopefully continue to collaborate with him. Ryan was another graduate student from Dr. Asher’s lab that I had the opportunity to work with. He’s a reliable, stand up fella, and never complained about riding the bus from Pitt down to Duquesne. Steve came along in my last year at Duquesne, but I think he is a smart genuine guy and hope to work with him further.

Lastly, but certainly not least, I’d like to acknowledge the great support, love, and encouragement offered by my family in the last six years. My mother, Kathi, is the most selfless soul you will ever meet, and she definitely thinks I am smarter and better looking than I really am. She is my number one fan and is confident I can accomplish great things, even when I am not so sure. My dad, Jim, is one of the most reasonable, intelligent, and undeniably antagonistic people I have ever encountered. He passed
on to me many great qualities such as: the desire to ask questions I already know
the answer to, a thirst for knowledge about the world around me, and an absolute
inability to ever do anything mean. My brothers, Scotty and Conner, have become
great friends of mine and outspoken supporters of my various endeavors, even if we
mostly just tease each other all the time. They’re great, believe me.

None of this would be possible without the funding and resources provided by
various organizations. I’d like to thank Duquesne University for employing me as a
lab assistant and graduate student, the Department of Chemistry and Biochemistry
for putting up with me for all this time, the Bayer School of Natural and Environ-
mental Sciences, and the Center for Computational Sciences for providing... well...
computational resources. Thanks to the University of Pittsburgh, which is where my
collaborators in the Dr. Asher group were located. Duquense’s Center for Computa-
tional Sciences was funded by the National Institute of Health, and supercomputing
resources were provided by the Extreme Science and Engineering Discovery Envi-
ronment (XSEDE), supported by National Science Foundation grant number ACI-
1548562. Also, thanks to the members of the Phi Lambda Upsilon (PLU) graduate
student society for planning annual research symposia on campus and organizing
educational and social events.

A final note, thanks to Emily Benner for helping me write this.
TABLE OF CONTENTS

Abstract .................................................................................................................... iv

Dedication .................................................................................................................. vi

Acknowledgements ................................................................................................... vii

List of Tables ............................................................................................................ xii

List of Figures .......................................................................................................... xiii

List of Abbreviations ............................................................................................... xiv

Chapter 1 INTRODUCTION TO POLYGLUTAMINE AGGREGATION RESEARCH AND MOLECULAR DYNAMICS .......................................................... 1

  1.1 Introduction ........................................................................................................ 1

  1.2 PolyQ Diseases .................................................................................................. 1

  1.3 Experimental Approaches ............................................................................... 3

  1.4 Computational Methods .................................................................................... 4

  1.5 Importance of Computational and Experimental Collaboration ............... 7

  1.6 Enhanced Sampling with Metadynamics ......................................................... 7

  1.7 Summary of Dissertation Work ....................................................................... 12

Chapter 2 STRUCTURAL ANALYSIS OF Q10 PEPTIDE FIBRIL POLYMORPHS .......................................................... 17

  2.1 Review of PolyQ Fibril Structure .................................................................... 17

  2.2 MD Simulations and UVRR Experiments on Q10 Fibrils ......................... 21

  2.3 Analysis of the Structural Properties of Q10 Fibril Polymorphs ............... 25

  2.4 Conclusions and Direction ............................................................................ 30

Chapter 3 CALCULATION AND EXPLORATION OF THE MONOMERIC CONFORMATIONAL ENERGY LANDSCAPE OF Q10 PEPTIDES ........................... 33

  3.1 Review of Aqueous PolyQ Monomeric Structure .......................................... 33

  3.2 MD Simulations and UVRR Spectroscopy of Aqueous Q10 Monomers ....... 35

  3.3 Characterization of the Monomeric Q10 Conformational Ensemble .......... 37

  3.4 Discussion and Conclusions ............................................................................ 47

Chapter 4 ANALYSIS OF RELATIVE HYDROGEN BONDING STRENGTHS FOR Q10 PEPTIDES AND FIBRILS ................................................................. 51

  4.1 Background on the Role of H-bonding in PolyQ Systems ............................ 51
4.2 Calculating the Interaction Enthalpy of Q10 H-bonding with MD and UVRR ............................................. 52
4.3 Evaluating the Comparative Strengths of Q10 Side Chain H-bonds ......................................................... 55
4.4 Conclusions and Direction ................................................................. 57

Chapter 5 LENGTH DEPENDENT PROPERTIES IN THE STRUCTURE OF Q15 AND Q20 PEPTIDE MONOMERS AND FIBRILS .................................................. 58
  5.1 Overview of Length Dependent Properties in PolyQ Peptide Structure .......................... 58
  5.2 Investigation of Q15 and Q20 monomeric structural properties with MD and UVRR .................................................. 59
  5.3 Determination of Length Dependent Structural Properties for Q10, Q15, and Q20 ........................................ 60
  5.4 Conclusions and Direction ................................................................. 68

Chapter 6 CONCLUSION ................................................................. 69
  6.1 Summary of Dissertation Work ................................................................. 69
  6.2 Future Directions and Unanswered Questions ........................................ 72

References ................................................................. 75
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>PolyQ Disease Types</td>
<td>2</td>
</tr>
<tr>
<td>3.1</td>
<td>Energy Analysis of PPII and 2.5\textsubscript{1}-helix MD Structures</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>H-bonding Strengths for Q10 Side Chains</td>
<td>55</td>
</tr>
<tr>
<td>5.1</td>
<td>Relative Free Energy of Q10, Q15, and Q20 Structural Minima</td>
<td>67</td>
</tr>
<tr>
<td>5.2</td>
<td>Transition energy barriers for Q10, Q15, Q20</td>
<td>67</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Sampling a Conformational Landscape at Equilibrium</td>
<td>9</td>
</tr>
<tr>
<td>1.2</td>
<td>Sampling a Conformational Landscape with Metadynamics</td>
<td>12</td>
</tr>
<tr>
<td>1.3</td>
<td>PolyQ Aggregation Mechanism Hypotheses</td>
<td>13</td>
</tr>
<tr>
<td>2.1</td>
<td>Schematic of polyQ aggregation</td>
<td>18</td>
</tr>
<tr>
<td>2.2</td>
<td>Side chain zipper</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Proposed Monomeric Structure in Fibrils</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>$\beta$-sheet geometries</td>
<td>20</td>
</tr>
<tr>
<td>2.5</td>
<td>Peptide Dihedral Angles</td>
<td>22</td>
</tr>
<tr>
<td>2.6</td>
<td>Q10 Fibril Models</td>
<td>23</td>
</tr>
<tr>
<td>2.7</td>
<td>Q10 Fibril MD Trajectories</td>
<td>26</td>
</tr>
<tr>
<td>2.8</td>
<td>Q10 Fibril RMSD vs. Time</td>
<td>26</td>
</tr>
<tr>
<td>2.9</td>
<td>Q10 Fibril $\Psi$ Angle Distributions</td>
<td>27</td>
</tr>
<tr>
<td>2.10</td>
<td>Q10 Fibril $\Psi$ Angle Distributions</td>
<td>30</td>
</tr>
<tr>
<td>2.11</td>
<td>Side Chain H-bonding in Q10 Fibrils</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>Comparison of $\Psi$ and $\chi_3$ Dihedral Angles for Q10 Monomers</td>
<td>38</td>
</tr>
<tr>
<td>3.2</td>
<td>Monomeric Q10 Conformational Energy Landscape</td>
<td>40</td>
</tr>
<tr>
<td>3.3</td>
<td>Ramachandran Plots for Q10 Monomeric Structures</td>
<td>41</td>
</tr>
<tr>
<td>3.4</td>
<td>H-bonding Populations for Q10 Monomers</td>
<td>43</td>
</tr>
<tr>
<td>3.5</td>
<td>Energy Barrier of Q10 Monomeric Energy Landscape</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>Q10 Monomeric and Fibril Structures for H-bonding Calculations</td>
<td>53</td>
</tr>
<tr>
<td>5.1</td>
<td>$\Psi$ Distributions for Q15 and Q20 Monomeric Structures</td>
<td>61</td>
</tr>
<tr>
<td>5.2</td>
<td>Monomeric Q15 Conformational Energy Landscape</td>
<td>63</td>
</tr>
<tr>
<td>5.3</td>
<td>Monomeric Q20 Conformational Energy Landscape</td>
<td>64</td>
</tr>
<tr>
<td>5.4</td>
<td>CD Spectra of Monomeric Q15 and Q20</td>
<td>66</td>
</tr>
</tbody>
</table>

xiii
LIST OF ABBREVIATIONS

PolyQ  Polyglutamine
H-bond  Hydrogen bond
MD  Molecular dynamics
UVRR  Ultraviolet resonance Raman
NMR  Nuclear magnetic resonance
EM  Electron microscopy
RMSD  Root mean square deviation
Q10  D_{2Q_{10}K_2}
Q15  D_{2Q_{15}K_2}
Q20  D_{2Q_{20}K_2}
PPII  Polyproline-II
VMD  Visual Molecular Dynamics
NAMD  Nanoscale Molecular Dynamics
vdW  van der Waals
Chapter 1. INTRODUCTION TO POLYGLUTAMINE AGGREGATION RESEARCH AND MOLECULAR DYNAMICS

1.1 Introduction

The study of polypeptide aggregation has become an area of intense interest following the discovery that the formation of proteinaceous aggregates is a hallmark of the molecular pathology of various neurodegenerative diseases, such as Alzheimer’s, Parkinson’s, and Huntington’s disease. In these diseases, the protein-rich aggregates share specific structural traits, such as high β-sheet content, fibrillar morphologies, and aqueous insolubility, and are collectively categorized as “amyloid” fibrils. The disease state for each of these diseases is linked to a different protein: β-amyloid, α-synuclein, and huntingtin for Alzheimer’s, Parkinson’s, and Huntington’s respectively. Although these proteins lack similar sequence or function, they aggregate into fibrils with remarkably similar features, including major β-sheet structure, binding of thioflavin-T, and insolubility. This property has led many to conclude that there must be significant similarities in the molecular pathology of these diseases. In this dissertation, peptide systems associated with one subclass of these diseases, polyglutamine (polyQ) diseases, which includes Huntington’s disease, are studied to better understand the molecular structures that play a role in the aggregation of polyQ peptides.

1.2 PolyQ Diseases

Aggregation of polyQ-rich polypeptides in neurons is a pathological marker of nine known neurodegenerative diseases. Each polyQ disease is caused by mutations that result in expanded CAG codon repeats in various genes, which consequently impart glutamine amino acid repeat expansions into the proteins encoded by the mutated genes. These elongated polyQ tracts cause the affected proteins to misfold in the cell and associate into peptide aggregates that mature into fibrils. The amount of glutamine repeats necessary to cause disease states differs per polyQ disease; each
disease possesses a critical value of repeats that elicits the disease.\textsuperscript{10,13} Table 1.1 shows the different associated proteins and critical repeat lengths for the nine polyQ diseases that have been identified.

<table>
<thead>
<tr>
<th>PolyQ disease</th>
<th>Associated protein</th>
<th>Critical repeat length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentatorubropallidoluysian atrophy</td>
<td>Atrophin-1</td>
<td>49</td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td>Huntingtin</td>
<td>36</td>
</tr>
<tr>
<td>Spinobulbar muscular atrophy</td>
<td>Androgen receptor</td>
<td>38</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 1</td>
<td>Ataxin-1</td>
<td>49</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 2</td>
<td>Ataxin-2</td>
<td>33</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 3</td>
<td>Ataxin-3</td>
<td>55</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 6</td>
<td>( CaV_{2.1} )</td>
<td>21</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 7</td>
<td>Ataxin-7</td>
<td>38</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 17</td>
<td>TBP</td>
<td>47</td>
</tr>
</tbody>
</table>

\textbf{Table 1.1: PolyQ disease types}\textsuperscript{14}

The fact that longer repeat lengths trigger the onset of polyQ diseases indicates that an increase in polyQ tract length speeds up the kinetics of the peptide aggregation process.\textsuperscript{15} The kinetics of polyQ and, in general, amyloid fibril formation occur via a nucleation and growth mechanism,\textsuperscript{6,16} although the molecular details of this mechanism are not well understood. It is thought that longer polyQ repeats hasten the occurrence of a nucleating event that allows fibril growth to take place.\textsuperscript{16–19} This fits with the disease pathology of polyQ diseases where affected individuals with longer polyQ repeat lengths generally show onset of the disease at younger ages.\textsuperscript{9–11}

It is important to note that the precise role that the phenomenon of polyQ peptide aggregation and fibril formation plays in the disease state has not been established. Initially, the consensus was that the physical presence of these large fibrils, also known as cellular inclusions, disrupted natural cellular function in a cytotoxic way.\textsuperscript{6,9,20} However, this hypothesis has been undermined by the observation that the presence of large inclusions actually extends the life of neurons by sequestering intracellular polyQ peptides.\textsuperscript{21} The prevalent hypothesis is now that mature cellular fibrils do not play a harmful role to cellular function; rather it has been reported that smaller, soluble
polyQ aggregates act as the cytotoxic agents in these diseases.\textsuperscript{18,22,23} There are varying theories as to why this is the case, however most studies correlate these species translocating into the nucleus with cytotoxicity.

1.3 Experimental Approaches

The study of polyQ aggregation, and indeed many other amyloid diseases, has benefited greatly from the fact that polyQ peptides aggregate to form fibrils in vitro that are structurally identical to those formed in vivo. This indicates that the aggregatory behavior of these peptides is not dependent on an intracellular environment and is instead a more general chemico-physical phenomenon that can be understood outside the constraints of a biological scope. A rich variety of experimental techniques have been utilized to probe the behavior of polyQ systems, including but not limited to circular dichroism (CD) spectroscopy,\textsuperscript{20,24,25} electron microscopy (EM),\textsuperscript{17,26–28} solid-state nuclear magnetic resonance (NMR) spectroscopy,\textsuperscript{29,30} infrared spectroscopy,\textsuperscript{31,32} UV resonance Raman (UVRR) spectroscopy,\textsuperscript{25,33–35} and X-ray diffraction,\textsuperscript{28,36,37} as well as numerous kinetic and thermodynamic biochemical assays.\textsuperscript{8,12,38} The results from many of these variant experimental studies have not given rise to a unified consensus as to the properties and behavior of these systems, but have instead have yielded a myriad of competing, fragmented theories of the aggregative behavior of polyQ peptides and fibrils.

Inconsistency in the conclusions drawn from the experimental study of polyQ systems is caused by the significant qualitative interpretation necessary to transform experimental measurements into detailed molecular understanding of the structural dynamics of polyQ peptides. For example, diffraction patterns must be translated into structural models consistent with the data,\textsuperscript{36,37} frequency shifts must be deciphered to identify the chemical phenomena at play,\textsuperscript{25,29,31} and at the end of the day, many experimental analysis methods are simply not incisive enough to yield the types of observations that will contribute greatly to the understanding of polyQ aggregation.
The dissonance of experimental conclusions in the field of polyQ aggregation has made clear the need for alternate methods of investigation that can accompany and sharpen experimental findings.

1.4 Computational Methods

In the years since the development of the field of polyQ aggregation, computational simulations, in particular those using molecular dynamics (MD), have become an increasingly prevalent tool in the study of biophysical systems.\textsuperscript{39,40} MD simulations allow researchers to probe systems similar to those prepared in the laboratory at the molecular scale, enabling examination at a finer level than most experimental techniques can offer. The method was originally reported in a communication from Los Alamos National Laboratory by Fermi et al. and later pioneered in the study of biological systems by Karplus, Levitt, and Warshel,\textsuperscript{41,42} as recognized in their 2013 collective Nobel Prize in chemistry. This dissertation makes significant use of MD simulations, and for that reason a short introduction of the physics of these simulations is discussed here.

MD makes use of Newton’s fundamental laws of motion\textsuperscript{40,43} to simulate the movements of the atoms that make up biomolecules. The forces at play in these atomistic systems are calculated with force fields, which are simply the equations and accompanying parameters used to calculate potential energy and force on the atoms.\textsuperscript{44} In modern MD, the velocity Verlet integration algorithm,\textsuperscript{45,46} shown in Equations 1.1 and 1.2, is used to calculate the trajectories of the atoms in the system.

\begin{align*}
\vec{q}(t + \delta t) &= \vec{q}(t) + \vec{v}\delta t + \frac{1}{2}\vec{a}(t)\delta t^2 \quad (1.1) \\
\vec{v}(t + \delta t) &= \vec{v}(t) + \frac{1}{2}\left(\vec{a}(t) + \vec{a}(t + \delta t)\right)\delta t \quad (1.2)
\end{align*}

In this algorithm, $\vec{q}(t)$ is atomic coordinates at time $t$, $\vec{v}(t)$ is atomic velocities at time $t$, $\vec{a}(t)$ is atomic acceleration at time $t$, and $\delta t$ is a small time increment known.
as the time step. As the algorithm shows, $\vec{q}$ and $\vec{v}$ can be obtained for some future time $t + \delta t$, but this is not the case for $\vec{a}$. To calculate $\vec{a}$ at a new configuration, Newton's second law of motion (Equation 1.3) is used to calculate $\vec{a}(t)$ from the force on the atom at time $t$.

$$\vec{a}(t) = \frac{\vec{F}(\vec{q}(t))}{m}$$ (1.3)

As indicated in Equation 1.3, the force $\vec{F}$ on an atom is a function of only the atomic coordinates $\vec{q}(t)$. Because of this, $\vec{F}$ can be calculated from any molecular configuration. In MD simulations, a force field is used to calculate force from the potential energy function of the system via the relationship shown below.

$$\vec{F}(\vec{q}(t)) = -\nabla U(\vec{q}(t)) \equiv -\left(\frac{\partial}{\partial x}U(\vec{q}(t)) + \frac{\partial}{\partial y}U(\vec{q}(t)) + \frac{\partial}{\partial z}U(\vec{q}(t))\right)$$ (1.4)

Confusingly, force fields are actually defined as the potential energy function, $U(\vec{q}(t))$, and accompanying parameters, and the gradient of this function (Equation 1.4) is used to calculate the force. The relationships contained in Equations 1.1, 1.2, 1.3, and 1.4 form the central mathematical machinery used to perform MD simulations.

The force field is incredibly important in the accuracy of MD simulations. In the last 25 years, common force fields such as CHARMM\textsuperscript{47} and Amber\textsuperscript{48} have been continually improved and refined. Force fields are, for the most part, empirically derived and, thus, many approximations and simplifications are made concerning the physics of the system.\textsuperscript{49} For example, in MD simulations electrons are not treated explicitly; instead atoms are treated as spheres that possess a predefined partial charge related to the dipoles and relative electronegativities of the molecule. Common potentials used to model the bonding, angular, and torsional energies in covalently bound molecules are shown in Equations 1.5, 1.6, and 1.7.\textsuperscript{44,48,50}
In Equations 1.5, 1.6, and 1.7, $U_{ij}^b$ is the bond stretching energy between atoms $i, j$, $U_{ijk}^a$ is the angular energy between atoms $i, j, k$, and $U_{ijkl}^d$ is the dihedral energy between $i, j, k, l$. $k_b, k_a$, and $k_d$ are force constants for bond, angle, and dihedral potentials, $r_b, \theta_a$, and $\delta$ are equilibrium values for bonds, angles, and dihedrals, and $r_{ij}, \theta_{ijk},$ and $n\phi_{ijkl}$ are the current values of each. These potential functions have the benefit of being simple and easily differentiable with respect to $x, y, z$, which allows calculation of force, but they are also ideal, symmetrical functions that neglect the effect of quantum mechanical phenomena such as anharmonicity and vibrational coupling.\textsuperscript{44,48,51} Equations 1.8 and 1.9 show the two primary forms of nonbonded atomic interaction potentials in common force fields: the Lennard-Jones\textsuperscript{52} (Equation 1.8) and Coulombic (Equation 1.9) potentials.

\begin{equation}
U_{ij}^b(r) = \frac{1}{2}k_b(r_{ij} - r_b)^2
\end{equation}

\begin{equation}
U_{ijk}^a(\theta) = \frac{1}{2}k_a(\theta_{ijk} - \theta_a)^2
\end{equation}

\begin{equation}
U_{ijkl}^d(\phi) = k_a[1 + \cos(n\phi_{ijkl} - \delta)]
\end{equation}

\begin{equation}
U_{ij}^{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]
\end{equation}

\begin{equation}
U_{ij}^{C}(r_{ij}) = \frac{1}{4\pi\varepsilon_0} \cdot \frac{q_i q_j}{r_{ij}}
\end{equation}

In both nonbonded potentials, $r_{ij}$ indicates the distance between atoms $i$ and $j$. In the Lennard-Jones equation, $\epsilon_{ij}$ and $\sigma_{ij}$ are respectively the strength of the attractive interaction between atoms $i$ and $j$ and the distance at which the energy is zero. For the Coulombic potential, $\varepsilon_0$ stands for the permittivity of vacuum, and $q_i$ and $q_j$ are the partial charges on atoms $i$ and $j$. These nonbonded potentials are also empirical approximations: $U_{LJ}$ models weak attractive and strong repulsive
forces such as London dispersion forces and volume overlap repulsion and $U_C$ models like-charge repulsion and opposite-charge attraction.

All molecules simulated with MD are chemically inert and only interact with other molecules via electrostatics and van der Waals forces; bonds cannot break or form. Responsible usage of MD requires the identification of the force field most suited to studying the system of interest.\textsuperscript{53,54} After the evaluation of several force fields\textsuperscript{55,56} for use in simulating polyQ peptides, the CHARMM36\textsuperscript{50} force field was chosen and is used for the most part in this dissertation.

1.5 Importance of Computational and Experimental Collaboration

Most biophysics experiments measure average properties of the sample of interest,\textsuperscript{57} but MD allows one to monitor the structure and dynamics of individual molecules in a system;\textsuperscript{39,49} although average properties of the sample may still be computed and compared to experimental data.\textsuperscript{58} The real power of computational techniques like MD lies in this potential for comparison to traditional experimental methods.\textsuperscript{57,59} The atomistic dynamic and structural information that MD can report is an excellent supplement to experimental studies; theoretical results can strengthen and expand experimental observations while corroboration with experiment can help validate computational models.\textsuperscript{39,46,60}

1.6 Enhanced Sampling with Metadynamics

Many biological processes involve the formation of energetically unfavorable, and thus transient, macromolecular structures. A majority of experiments cannot probe these processes because analyses that respond to the average properties of a sample by their nature cannot monitor short-lived, rare events. However, MD simulations are able to probe these types of events robustly with a variety of computational techniques known as enhanced sampling methods, including conformational flooding,\textsuperscript{61} replica-exchange molecular dynamics,\textsuperscript{62} free-energy perturbation,\textsuperscript{63} accelerated MD,\textsuperscript{64} and umbrella sampling.\textsuperscript{65} The work presented in this thesis utilizes one such method.
called metadynamics,\textsuperscript{66} based on the concept of conformational flooding, which enables the efficient exploration and characterization of the structural energy landscape of biomolecules.

Two important aspects of MD simulations are fundamentally limited. The first deficiency is the inaccuracy of force fields, which has already been discussed. Secondly, MD simulations are held back by their high computational cost. Computation of force in a typical MD simulation requires the evaluation of at least $\sim N^{N-1}$ interatomic potentials, which contain costly square root calculations.\textsuperscript{67} Consequently, MD simulations require considerable resources to carry out. For example,\textsuperscript{68} 1\,\mu s of MD simulation on a 25,000 atom system with 24 processors will take weeks or even months of computation time and to simulate larger systems for this amount of time requires petascale supercomputers to run quickly.\textsuperscript{69}

This computing limitation often leads to inadequate sampling of conformational states, disabling the simulation from giving insight into the nature of the system of interest.\textsuperscript{67} In statistical mechanics, sampling refers to the amount of configurations, or arrangements of atoms in a system, that can be accessed. Every atomic arrangement has an associated free energy cost, and high free energy costs can require long time scales to access. Groups of similar configurations are denoted as “states”, and each landscape has different states with varying free energies. A conformational free energy landscape with abundant deep energy states (wells) and high energy states (barriers) is referred to as a “rugged” or “frustrated” landscape. Systems impeded by inadequate simulation time and rugged free energy landscapes may become trapped in deep wells at equilibrium, with insufficient thermal energy to traverse the associated barriers.\textsuperscript{69}

Figure 1.1 depicts such a free energy landscape for an example biomolecule, wherein the $x-$axis represents some measure of conformation (e.g. RMSD, dihedral angle) and the $y-$axis represents the free energy cost to reach a certain conformation on the landscape. The only way to enable a system such as the one shown in Figure 1.1
Figure 1.1: Example conformational free energy landscape for a biomolecule. Two deep energy wells (A and B) exist, but due to the temperature of the system, the molecule may only climb so high up the energy barrier (dotted red line). Unfortunately, at the time scale available for simulation, this system may not be able to sample well B. Source: figure prepared with Python package Matplotlib and Adobe Illustrator.

To sample state B is to raise the temperature high enough so the free energy of the dotted red line in Figure 1.1 exceeds the height of free energy barrier. However, often this is undesirable, as an elevated temperature causes the atoms to possess greater velocities, potentially invalidating experimental comparison at room temperature, for example. Although some enhanced sampling methods, such as replica-exchange MD,\textsuperscript{62} do employ temperature jumps, temperatures are lowered back down after some simulation time to allow new conformations to sample at reasonable conditions.

There are three main types of enhanced sampling methods. Methods like umbrella sampling\textsuperscript{65} and free energy perturbation\textsuperscript{63} split the configurational space into subsections called “windows” and multiple simulations are run with one simulation sampling one window. Force constants\textsuperscript{65} or scaling parameters\textsuperscript{63} are used to restrict
the sampling of each simulation to the associated window. This type of enhanced sampling is robust if enough windows are able to be sampled, however if the windows are too large the sampling can be insufficient.\textsuperscript{69} Also, these methods are not useful for sampling an \textit{entire} energy landscape for a complex system, as the number of required simulation windows would be prohibitive.\textsuperscript{65}

Another type, which was discussed briefly two paragraphs prior, are temperature modulation methods such as replica-exchange MD.\textsuperscript{62} Sampling with these methods involves temporarily boosting the temperature of the system, allowing it to reach states that would be inaccessible at lower temperatures. These methods are well-suited for speculative conformational searching, but lack the quantitative ability to calculate the shape of the free energy landscape.\textsuperscript{66,67}

The third type of enhanced sampling involves use of an alternate potential energy function for the system that encourages efficient sampling of the conformational free energy landscape. This family encompasses methods like metadynamics,\textsuperscript{66} conformational flooding,\textsuperscript{61} and accelerated MD.\textsuperscript{64} These methods are particularly flexible because they do not require the sampling of discrete windows, and, in the case of conformational flooding\textsuperscript{61} and metadynamics,\textsuperscript{66} allow the quantification of the sampled free energy landscape.

The way methods like conformational flooding\textsuperscript{61} and metadynamics\textsuperscript{66} improve sampling of frustrated free energy landscapes is by adding \textit{artificial} energy to the potential energy function at that causes deep energy wells to become more shallow. After all rugged energy wells and barriers are sampled, the conformational energy landscape flattens out and possesses equal probability states. This added energy allows the system to efficiently escape energy minimi and traverse energy maxima while still sampling according to the physics of the underlying free energy landscape.

In metadynamics,\textsuperscript{66} artificial energy is added to the potential energy of the system as unnormalized Gaussian functions, or “hills”, the form of which is given in
Equation 1.10.

\[ U_{\text{Gauss}}(\xi) = W \exp \left( -\frac{(\xi(t) - \xi(t - \delta t))^2}{2\sigma^2} \right) \]  

(1.10)

Here, \( U_{\text{Gauss}} \) is the energy of one added Gaussian hill, \( W \) is the coefficient that determines the height of the hill, \( \sigma \) represents the standard deviation of the Gaussian, and \( \xi(t) \) and \( \xi(t - \delta t) \) are the positions on the conformational energy landscape at times \( t \) and \( t + \delta t \). Conformational energy landscapes are constructed using coordinates that simplify three-dimensional space into “collective variables” \(^{43,67,69}\) Collective variables allow the construction of conformational space using simple measures of conformation that are functions of Cartesian coordinates, such as RMSD, radius of gyration, and dihedral angles. In the metadynamics simulations presented herein, \( \xi \) will be RMSD values with respect to the \( \alpha \)-carbons of ideal secondary structures.

Values for \( W \) and \( \sigma \) in Equation 1.10 must be chosen with scrutiny: \( W \) must be maximized while being less than the average force on the unbiased system, and \( \sigma \) should be between \( \frac{1}{20} \) and \( \frac{1}{10} \) of the length of the full conformational space.\(^70\) Laio et al.\(^70\) discovered the estimated length of a metadynamics simulation can be formulated as follows

\[ t_{\text{total}} = \frac{\Delta \bar{G}_{\text{est}}}{W} \tau_{\text{Gauss}} \left( \frac{\Delta \bar{\xi}_{\text{est}}}{\sigma} \right)^d \]  

(1.11)

where \( \Delta \bar{G}_{\text{est}} \) is the estimate of the highest free energy barrier to traverse, \( W \) is the same from Equation 1.10, \( \tau_{\text{Gauss}} \) is the frequency in MD steps that Gaussian hills are deposited, \( \Delta \bar{\xi} \) is the total width of the collective variable space, \( \sigma \) is the same from Equation 1.10, and \( d \) is the number of dimensions that make up the collective variable space. In the metadynamics simulations discussed in this dissertation, \( d = 3 \).

Figure 1.2 shows the same landscape depicted in Figure 1.1 but after metadynamics has been used to sample energy well \( A \). The added Gaussian energy cause the potential energy of system to be high enough that it can easily mount the barrier.
Figure 1.2: Example conformational free energy landscape for a biomolecule. Two deep energy wells \((A\) and \(B\)) exist, and metadynamics sampling has been used to add artificial energy (green hills) into energy well \(A\) to allow the system to climb the barrier separating wells \(A\) and \(B\). Source: figure prepared with Python package Matplotlib and Adobe Illustrator.

and sample into well \(B\). A useful property of metadynamics is that, with the history of deposited Gaussian hills in a simulation, one can recreate the conformational free energy landscape sampled by the system during the simulation.\(^{66,70}\) This will be used in this dissertation to explore and characterize the free energy landscape of polyglutamine peptides.

1.7 Summary of Dissertation Work

Our goal in this dissertation was to better understand the structural properties of polyQ peptides in various states that are present along the aggregation mechanism. The mechanism by which polyQ peptides aggregate and eventually form insoluble amyloid fibrils is not well understood. Understanding this mechanism is important because polyQ diseases are reported to be primarily gain-of-function diseases.\(^{18,71,72}\)
As a consequence, there is great interest in characterizing the molecular mechanisms at play in these diseases. Figure 1.3 shows a hypothetical polyQ aggregation mechanism\textsuperscript{25,27,73} constructed from various theories explaining the aggregation of polyQ peptides.

**Figure 1.3:** Flowchart diagram illustrating various hypotheses for polyQ peptide aggregation. The first row of the flowchart illustrates possible monomeric structures for polyQ peptides that can induce various types of peptide dimerization events. Posited dimer structures are shown in the middle row, leading to ordered aggregate structures in the third row. *Source:* figure prepared with Adobe Illustrator.

There are two main theories for the molecular mechanism of polyQ aggregation. One theory,\textsuperscript{8,12,13,27,30,37,74} called here “nucleation and growth”, states that polyQ monomers come together to form an ordered $\beta$-hairpin or $\beta$-sheet oligomer, and monomers are added on to this initial structure, allowing the aggregate to grow (shown in the right and left pathways of Figure 1.3). Alternatively,\textsuperscript{19,73,75–78} the middle section of Figure 1.3 shows another pathway, named here “disordered cluster”, where polyQ monomers associate into disordered clusters of peptides that eventually convert to the structurally ordered aggregates seen on the bottom right and left of the figure.

Accurately determining the molecular mechanism of aggregation for polyQ peptides is key, because that knowledge can inform the search for treatment strategies.
for polyQ diseases. Several studies have reported molecules that are able to interfere with the aggregation of the polyQ tract of the huntingtin protein, such as Congo red, thioflavin S, and trehalose. However, due to lack of knowledge of the aggregation mechanism it is unclear what polyQ species these molecules bind to or inhibit. A complete picture of the aggregation mechanism could give researchers the clarity required to identify specific molecular targets for inhibitors like these. However, this picture must be built piecewise. Current simulation methods lack the computing power to simulate the aggregation of polyQ fibrils en masse. Instead, an understanding of specific steps of the flowchart in Figure 1.3 can be developed and compiled into a singular theory for polyQ aggregation.

This dissertation contains work that deepens the understanding of the polyQ aggregation mechanism by determining the structure of polyQ fibrils. As shown in the bottom row of Figure 1.3, the field is unsure of the structural motif of polyQ fibrillar aggregates: some reporting structures composed of β-hairpins and others positing β-sheet fibrils formed extended β-strands. Understanding the structural makeup of polyQ fibrils is essential to understanding the aggregation process, seeing as fibrils are the endpoint of the mechanism.

The second chapter of this dissertation aimed to determine the structural propensity of polyQ fibrils formed from short polyQ peptides. The structures of polyQ fibrils composed of $D_2Q_{10}K_2$ (Q10) peptides were investigated using classical MD simulations and UVRR experiments. Three polymorphs of Q10 fibrils were simulated and data was compared to UVRR results, showing that Q10 primarily adopts primarily antiparallel extended β-sheet fibril structures while able to adopt minor populations of parallel extended β-sheets. These results indicate that, at least for short polyQ sequences, the β-hairpin-based aggregation hypothesis shown as the right pathway in Figure 1.3 is not the correct view. Rather, an aggregation mechanism ending with extended antiparallel β-sheet fibrils (leftmost pathway in Figure 1.3) should be de-
In the third and fifth chapters of the dissertation, the first row of Figure 1.3, monomeric polyQ structure, was investigated. In Chapter 3, we used metadynamics MD simulations to characterize the conformational free energy landscape of aqueous, monomeric Q10 peptides. Results from these simulations were compared with data from UVRR spectra of Q10, revealing that monomeric Q10 can adopt stable collapsed $\beta$-strand and PPII-rich secondary structures. We observed a prohibitive energy barrier separating these structures from a $\beta$-hairpin conformation, and thus find it unlikely that Q10 forms such a structure. This observation supports our findings from the first chapter that $\beta$-hairpins are not present on the Q10 aggregation mechanism.

The fifth chapter contained a study in which we probed the monomeric structural ensemble of longer $D_2Q_{15}K_2$ (Q15) and $D_2Q_{20}K_2$ (Q20) peptides to observe the effect of increased repeat length on the structure of polyQ monomers. This is an important case because it can forecast how the aggregation mechanisms depicted in Figure 1.3 would change given different polyQ sequences. The results of this study indicate that increased polyQ length decreases the energy barrier separating $\beta$-strand and PPII-rich secondary structures. This phenomenon may explain the earlier onset of polyQ diseases that occurs with longer repeat lengths. It also indicates that monomeric $\beta$-hairpins are still not present in longer polyQ sequences and likely do not play a role in the early stages of aggregation.

Chapter 4 describes a study in which we investigated, in detail, the strengths of various types of Q10 hydrogen bonding (H-bonding). Glutamine H-bonding is thought to play an important role in the monomeric, oligomeric, and fibril structures in the polyQ aggregation mechanism, and here MD simulations and UVRR demonstrated that Q10 forms stronger intrapeptide H-bonds than those formed between the peptide and water. This result strongly indicates that polyQ H-bonding is a driving force in
the association of polyQ peptides, and points toward a possible strategy of disrupting polyQ H-bonding to interfere with polyQ aggregation.

The goal of this dissertation was to provide insight into polyQ structural states that exist along the aggregation pathway, and by doing so, pave the way for future development and wholistic understanding of treatment options.
Chapter 2. STRUCTURAL ANALYSIS OF Q10 PEPTIDE FIBRIL POLYMORPHS

This section is partially adapted and reproduced (with permission) from the following manuscript: Punihaole, D.; Workman, R.J.; Hong, Z.; Madura, J.D.; Asher, S.A. The Journal of Physical Chemistry B. 2016, 120(12), 3012-3026.

2.1 Review of PolyQ Fibril Structure

It is clear that polyQ-rich aggregates play an important role in the pathology of polyQ diseases such as Huntington’s.6,10,11,72 Given this fact, it is key to understand the structure of polyQ-rich aggregates; information that may aid researchers in inhibiting or reversing the aggregation process. A simple diagram of polyQ fibril formation is depicted in Figure 2.1.

Numerous studies of the structure of polyQ fibrils have been published using a plethora of biophysical analytical methods.8,12,17,20,24–38 In 1994, Perutz et al.20 reported one of the earliest investigations of polyQ fibril structure, examining the structure of D$_2$Q$_{15}$K$_2$ peptide fibrils. The aspartate and lysine amino acid pairs at the termini of this peptide enable solubility in water. Residues with charged side chains are now commonly appended to glutamine repeat peptides for this reason.

Perutz and coworkers observed peptide aggregates with a wide range of molecular weights with size exclusion chromatography assays.20 They used circular dichroism spectroscopy (CD) to probe the secondary structure of solutions of the D$_2$Q$_{15}$K$_2$ peptides. CD spectra for these samples were uniformly indicative of β-sheets; this was the first out of many studies that would confirm that polyQ peptides form β-sheet aggregates and fibrils that are similar to those formed in amyloid diseases such as Alzheimer’s and Parkinson’s.6,20,31,36,37

X-ray diffraction studies have shown that polyQ fibrils of various glutamine repeat lengths all possess similar diffraction patterns.36,37 However, the authors of these
Figure 2.1: Simple schematic of polyQ aggregation. Aqueous monomeric peptides associate to form ordered aggregates which eventually form mature insoluble fibrils in the neuron (shown in an electron micrograph). Source: figure prepared with Adobe Illustrator.

reports have assigned very different structures from relatively similar diffraction patterns. Perutz et al.\textsuperscript{20,28} concluded that X-ray diffraction data for a variety of amyloid fibrils, D\textsubscript{2}Q\textsubscript{15}K\textsubscript{2} fibrils included, are consistent with water-filled $\beta$-nanotubes stabilized by side chain interdigitation, a phenomenon that has been termed a “polar” or “steric” zipper. Zipper interactions in a pair of stacked $\beta$-sheets are shown in Figure 2.2.

Figure 2.2: $\beta$-sheets forming side chain zipper. Source: figure prepared with VMD.\textsuperscript{83}

Sikorski et al.\textsuperscript{36} analyzed the structure of D\textsubscript{2}Q\textsubscript{15}K\textsubscript{2} fibrils and disagreed with
Perutz et al., concluding that these fibrils are composed of antiparallel $\beta$-sheets with a $\beta$-hairpin as the monomeric unit. Sharma et al., in an X-ray diffraction study of $Q_8$, $Q_{15}$, $Q_{28}$, and $Q_{45}$ fibrils, found that all four peptides formed slab-like antiparallel $\beta$-sheets made up of $\beta$-hairpins as opposed to $\beta$-nanotubes, agreeing with Sikorski's conclusion.

A few more recent structural analyses of polyQ fibril aggregates were carried out with solid-state NMR spectroscopy. One such study found that polyQ fibrils composed of physiologically relevant glutamine repeat lengths, i.e. $Q_{N>37}$, form an antiparallel $\beta$-sheet structure dubbed a “$\beta$-arc”. Example $\beta$-arc and $\beta$-hairpin peptide structures are depicted in Figure 2.3.

![Figure 2.3: Two polyQ structures proposed to make up fibrils: (a) $\beta$-arc and (b) $\beta$-hairpin. Source: figure prepared with VMD.](image)

In $\beta$-arc fibril structures, individual peptides occupy multiple layers of $\beta$-sheets, similar to fibril structures for the Alzheimer’s A$\beta$ protein reported by Petkova and others. It is worth noting that all other instances of $\beta$-arc fibril structures were observed to be parallel $\beta$-sheets, which is not the case in polyQ studies. Geometrical differences between antiparallel and parallel $\beta$-sheets are depicted in Figure 2.4.
Antiparallel $\beta$-sheets (shown on the left in Figure 2.4) form linear amide-amide H-bonds, where parallel $\beta$-sheets (right side of Figure 2.4) form staggered amide-amide H-bonds. In antiparallel $\beta$-sheets, the $\Psi, \Phi$ dihedral angles of the backbone cause the structure to be slightly lower in energy.$^{85}$

![Antiparallel $\beta$-sheet](image1) ![Parallel $\beta$-sheet](image2)

**Figure 2.4:** Geometries of antiparallel and parallel $\beta$-sheets. Average $(\Phi, \Psi)$ angles for each type of $\beta$-sheet are shown. *Source:* figure prepared with VMD$^{83}$ and Adobe Illustrator.

However, other solid-state NMR work suggests alternative structural motifs; Sivanan-dam et al.$^{86}$ observed NMR data consistent with $\beta$-sheets formed of $\beta$-hairpin monomers, and kinetics work by Kar et al.$^{13,30}$ asserts that $\beta$-hairpin turns are critical to the aggregation kinetics of Q$_{23}$ and Q$_{26}$ peptides, although this does not directly correspond to the presence of $\beta$-hairpins in large fibril aggregates.

MD simulation approaches have been utilized in recent years to investigate the structural properties of polyQ-rich fibrils$^{[23, 82]}$. Most of these computational studies probe the kinetic or thermodynamic stability of different types of fibril structures posited by past experimental observations. Work by Esposito et al.$^{23}$ found that Q$_{15}$ is stable in extended, antiparallel $\beta$-sheets and side chain zippers play an important role in stabilizing, but not forming, the fibrils. Another study$^{82}$ found that among $\beta$-sheet motifs for short polyQ peptides, antiparallel $\beta$-sheets form more stable fibrils than parallel $\beta$-sheets. However, the majority of these studies are carried out independent of experimental considerations, and as a result direct comparison of computational and experimental findings is rare.
Overall, there is a lack of consensus regarding the structural properties of polyQ peptide fibrils; this serves to underscore the need for incisive and reliable biophysical methods that can discriminate between the myriad of proposed polyQ fibril models in a quantitative, clear way. One main component of understanding the structural behavior of polyQ fibrils is characterizing the conformations and H-bonding interactions of glutamine side chains, which have been reported by many of the aforementioned polyQ studies to play a crucial role in the stability and selectivity of polyQ structures. There is also uncertainty in the field as to the type of $\beta$-sheets that these peptides form.$^{20,29,36,37,87}$ As mentioned earlier, other amyloidogenic fibrils form primarily parallel $\beta$-sheets.$^6,84$ However, there have been numerous reports of polyQ fibrils forming antiparallel $\beta$-sheets instead.$^{20,36,37}$ This is an important distinction, because currently, polyQ fibrils are categorized under the larger umbrella of amyloid fibrils. If the structure of polyQ fibrils is fundamentally different than these other types of fibrils in the type of $\beta$-sheet that forms, this may serve as an important distinction between the groups of fibril diseases.

2.2 MD Simulations and UVRR Experiments on Q10 Fibrils

In this work, MD simulations and UVRR experiments are coupled to determine the structure properties of polyQ fibrils prepared from D$_2$Q$_{10}$K$_2$ (Q10) peptides. Previously, Xiong et al.$^{25}$ observed that this peptide can adopt two conformational states in aqueous solution, a hypothesized $\beta$-hairpin-like structure prepared from powder peptide samples (called NDQ10) and a PPII-rich conformation prepared by disaggregating Q10 fibrils (called DQ10). Both of these structural states are shown to aggregate into amyloid-like fibrils.

UVRR spectroscopy is used here to measure the backbone $\Psi$ dihedral angle distributions of fibril solutions formed from the NDQ10 and DQ10 monomeric states, as well as the $\chi_3$ side chain dihedral angles. Figure 2.5 shows a visual description of these two dihedral angles, as well as the $\Phi$ dihedral angle. Distributions of these
dihedral angles are quantitatively compared to $\Psi$ and $\chi_3$ angle distributions gathered from MD simulations of three fibril models.

The three Q10 fibril models (shown in Figure 2.6) simulated in this work were an antiparallel extended $\beta$-sheet (Model $a$), a parallel extended $\beta$-sheet (Model $b$), and an antiparallel $\beta$-hairpin $\beta$-sheet (Model $c$). These systems contained eight Q10 peptides that were assembled into parallel or antiparallel fibril structures. To begin, two single $\beta$-sheets made up of four peptides were constructed. One $\beta$-sheet was then flipped $180^\circ$ so that each strand was oriented in the opposite direction of the peptides in the other $\beta$-sheet. The flipped $\beta$-sheet was then stacked on top of the other sheet to form a stacked $\beta$-sheet structure. Rotation of one $\beta$-sheet was necessary to ensure the positively charged termini of one $\beta$-sheet were aligned with the negatively charged termini of the other. Canonical ($\Phi$, $\Psi$) angles of ($-140^\circ$, $135^\circ$) were used to create the antiparallel $\beta$-strands in Model $a$, and angles of ($-120^\circ$, $113^\circ$) were used to make the parallel strands in Model $b$. Model $c$ was made using $\beta$-hairpin geometries that had been observed in previous unpublished work by the Madura group. All model fibrils were constructed with the Molecular Operating Environment (MOE) software suite and Visual Molecular Dynamics (VMD).

Each eight-peptide fibril model was solvated in a periodic $70 \times 50 \times 50 \, \text{\AA}^3$ water
Figure 2.6: Octameric Q10 fibril models simulated in Chapter 2. Model \( a \) is an antiparallel extended \( \beta \)-sheet, Model \( b \) an extended parallel \( \beta \)-sheet, and Model \( c \) a \( \beta \)-hairpin-sheet. Each model composed of two stacked \( \beta \)-sheets with four peptides per sheet. Source: figure prepared with Chimera and Adobe Illustrator.

box of 5087 water molecules and energy minimized for 10,000 steps using the conjugate gradient minimization method.\(^{43}\) The three systems were then equilibrated at
constant pressure and temperature for 50 ps. During this minimization and equi-
libration, the $x, y, z$ positions of all the protein atoms of the fibrils were restrained by
a harmonic $\frac{1}{2}k(x - x_0)^2$ potential. Following this equilibration, restraints were ap-
plied only to the positions of the backbone (O–C–C$_\alpha$–N) atoms of the peptides, and
another 10 000 steps of minimization and 100 ps of equilibration at constant pressure
and temperature were carried out. The aim of this protocol of restrained minimiza-
tion and equilibration is to allow the water in the system to find the lowest energy
configuration around the fibril structures before the peptides are allowed to freely
move. After this protocol was completed for each system, 50 ns of completely unre-
strained equilibration (no harmonic constraints) was simulated, followed by 200 ns of
data production MD.

The NAMD$^{43}$ software package (Version 2.10) was used to simulate MD in this
work. Potential energies and forces were obtained with the CHARMM36$^{50}$ force field.
CHARMM36 was selected for its torsional energy corrections intended to decrease
an $\alpha$-helix bias previously associated with CHARMM and properly stabilize $\beta$-strand
conformations.$^{89}$ Alternate force fields such as Amber99ffSB$^{56}$ also include the CMAP
corrections, but CHARMM22 was chosen due to its reported ability to describe $\alpha$-
helical, $\beta$-sheet, and random coil structures accurately.$^{89}$

The TIP3P$^{90}$ water model was used to simulate water in all simulations, along
with the particle mesh Ewald algorithm$^{91}$ for full-system electrostatics with a grid
spacing of 1.0 Å. The Verlet velocity algorithm$^{45}$ was used to integrate the equations
of motion with a time step of 2 fs. All simulations were performed under constant
pressure and temperature (NPT) with a Langevin thermostat and piston utilized to
maintain a temperature of 300 K and pressure of 1.01325 bar respectively.$^{43,92}$ A pair
interaction cutoff of 12.0 Å was used with a nonbonded switching distance of 10.0 Å.

Analysis was performed with Visual Molecular Dynamics$^{83}$ (VMD) and its native
Tcl scripting. $\Psi$ and $\chi_3$ dihedral angles were collected from all glutamine residues with
a Tcl script in VMD.\textsuperscript{83} $\chi_3$ side chain dihedral angles were only taken from glutamine side chains that were in the buried interior of the fibril models, to best simulate the interior of a bulk fibril environment. Extent of dissociation for the fibril models was determined by visual inspection of the structures along with an RMSD metric for the fibril structure calculated with respect to the backbone atoms of the peptides. The Gibbs free energy difference, $\Delta G$, between Models $a$ and $b$ was calculated using a Python implementation of the Bennett acceptance ratio called Pymbar.\textsuperscript{93} Potential energies were extracted from NAMD\textsuperscript{43} log files and used as input for Pymbar.

H-bonding analysis was performed on Models $a$ and $b$ using VMD.\textsuperscript{83} Positive H-bond contacts were specified as a heavy atom (N-O) distance of $<3.0$ Å and an N-H-O angle between $-30^\circ$ and $30^\circ$. Backbone-backbone, backbone-side chain, side chain-side chain, and peptide-water H-bond populations were calculated over the course of each 200 ns trajectory. Histograms of these data were prepared with the statistical computing package R.\textsuperscript{94}

2.3 Analysis of the Structural Properties of Q10 Fibril Polymorphs

The 200 ns MD simulations of the computational fibril Models $a$, $b$, and $c$ indicate that the extended $\beta$-sheet fibrils, Models $a$ and $b$, are significantly more stable than the $\beta$-hairpin composed fibril. Figure 2.7a-c shows the structure of each fibril model at 50 ns intervals throughout the simulations. Panel 2.7c shows the dissociation of the $\beta$-hairpin fibril model, which occurs at roughly 60 ns into the 200 ns production simulation. Models $a$ and $b$, depicted in 2.5a,b, remain structurally intact throughout the entire simulation. The RMSD of the backbone atoms of each fibril model was recorded with respect to the initial structure; Figure 2.8 shows the progression of the RMSD traces for each model. An RMSD value of 3.0 Å was defined as the cutoff for dissociation, as visually the fibril structure of the $\beta$-hairpin fibril began to break apart above this level.

The results from these simulations indicate that Q10 does not form stable $\beta$-
Figure 2.7: Snapshots from the 200 ns MD trajectory of three 8-mer Q10 fibrils. Structures were solvated but water was not shown for purposes of illustration. Source: figure prepared with Chimera and Adobe Illustrator.

Figure 2.8: Plot of RMSD vs. time for the three Q10 fibril polymorphs. RMSD taken with respect to α-carbons of initial structure. Source: figure prepared with Python package Matplotlib.

To couple computational and experimental observations of the structure of Q10 fibrils, backbone Ψ dihedral angle distributions were calculated from each method.
and compared and contrasted. Figure 2.9 shows the Ψ distributions from both the experiments and computations.

![Figure 2.9](image_url)

**Figure 2.9**: Comparison of Ψ dihedral angle distributions for three Q10 fibril polymorphs from MD simulations and UVRR spectra. Canonical locations of parallel and antiparallel β-strand structure shown with dotted red line. Source: figure prepared with Python package Matplotlib and Adobe Illustrator.

Panels 2.9a,b show UVRR data from fibrils prepared from the NDQ10 and DQ10 monomeric conformational states. Both distributions feature prominent peaks at 145° in the antiparallel β-strand region, as well as a smaller shoulder located at around 125°, which corresponds to a minor population of parallel β-strand. The nearly identical nature of these distributions indicates that both NDQ10 and DQ10 form the same fibril structure. In Panels 2.9c-e Ψ distributions from the data production MD simulations of computational fibril Models a, b, and c are displayed. Distributions for Models a and b have large peaks at ~141° and ~127° respectively, which closely correspond to the locations of the experimentally observed peaks of 145° and 125° for antiparallel and parallel β-sheet populations. Model c yielded a distribution pos-
sessing a doublet at $\sim -19^\circ$ and $\sim -43^\circ$ that is not observed in the experimental distributions. This feature is indicative of the $i + 1$ and $i + 2$ residues of the type I $\beta$-turn present in the $\beta$-hairpins of Model c. The agreement between the $\beta$ distributions for Models a and b and those obtained from UVRR supports the conclusion that the structural motif for NDQ10 and DQ10 peptide fibrils are stacked $\beta$-sheets made up of extended $\beta$-strands.

Moreover, these models are consistent with the results from other polyQ work. Schneider et al.$^{29}$ posited, based on EM and solid-state NMR data, that D$_2$Q$_{15}$K$_2$ peptide fibrils consist of extended $\beta$-strands. Another study by Thakur and others$^{27}$ found that polyQ peptides with $\beta$-hairpin-inducing proline-glycine moieties only form stable $\beta$-sheet fibrils when each side of the $\beta$-hairpin contains stretches of at least 9-10 glutamine residues. The D$_2$Q$_{10}$K$_2$ peptides studied here would not be capable of forming $\beta$-hairpin fibrils based on that criterion. This strengthens our conclusion that Q10 fibrils are composed of extended strands rather than hairpins.

The next question we looked to address was whether Q10 adopts parallel or antiparallel $\beta$-sheets. Most amyloid fibrils form parallel $\beta$-sheets, which seem to maximize hydrophobic and steric zipper interactions. PolyQ fibrils are relatively unique among amyloid fibrils if they form antiparallel $\beta$-sheets, which many studies suggest they do.$^{36,37}$ To examine the preference between antiparallel and parallel $\beta$-sheets in polyQ fibrils, we used the Bennett acceptance ratio method$^{93}$ to calculate the free energy difference between fibril Models a and b. The free energy of the antiparallel $\beta$-sheet fibril system, Model a, was found to be $160.5 \pm 2.0$ kJ mol$^{-1}$ lower than the parallel $\beta$-sheet model, Model b. Because the fibrils were composed of eight Q10 peptides, we were able to infer that the free energy difference between Models a and b was $\sim 1.5$ kJ mol$^{-1}$ per peptide bond. These results provide further confirmation that antiparallel $\beta$-sheet formation is favored in polyQ fibrils and accounts for the greater fraction of Q10 antiparallel $\beta$-sheet observed in the experimental results. Our
collaborators were able integrate the amide vibrational bands in the UVRR spectra to estimate an analogous free energy difference per peptide bond for antiparallel and parallel Q10 fibrils. This yielded a free energy difference per peptide bond of \( \sim 1.0 \text{ kJ mol}^{-1} \), which corresponds well to the value obtained from simulation.

We investigated the electrostatic and H-bonding interactions of Models \( a \) and \( b \) to probe the cause of the energy favorability of antiparallel over parallel polyQ \( \beta \)-sheets. Electrostatically, Models \( a \) and \( b \) possess favorable interactions between the N-terminal aspartate residues and C-terminal lysine residues from opposing stacked \( \beta \)-sheets. Within the \( \beta \)-sheet, we found that Model \( a \) features favorable inter-sheet Coulombic interactions between oppositely charged terminal residues. However, the parallel \( \beta \)-sheet of Model \( b \) does not allow these inter-sheet opposite charge terminal interactions and like-charged terminal interactions destabilizes the fibril. This repulsion between the terminal residues in parallel Q10 \( \beta \)-sheets may perturb local backbone-backbone H-bonding.

Another quantitative marker we used to compare our MD simulations to our collaborator’s experiment in this work was the \( \chi_3 \) side chain dihedral angle \( (\text{C}_\beta-\text{C}_\gamma-\text{C}_\delta-\text{O}_\epsilon) \). Distributions for \( \chi_3 \) can be calculated from UVRR spectra in the same way as the \( \Psi \) angle because the side chain amide possesses the same structural sensitivity as the backbone amide.

Figure 2.10 is organized the same way as Figure 2.9, showing \( \chi_3 \) dihedral angle distributions for the NDQ10 and DQ10 structural states in Panels 2.10a and b, and distributions from the simulations of Models \( a \), \( b \), and \( c \) in Panels c-e. The experimental \( \chi_3 \) distributions in Panels a and b are similar, featuring two Gaussian populations of angles centered around \(-14^\circ\) and \(5^\circ\) for NDQ10 and \(-12^\circ\) and \(3^\circ\) for DQ10. For the computational results, the antiparallel \( \beta \)-sheet Model \( a \) has a salient peak at \(4^\circ\) and the parallel \( \beta \)-sheet Model \( b \) gives a distribution with a dominant peak centered at \(-10^\circ\). Each of these peaks correspond well to the distributions obtained from the
Figure 2.10: Comparison of \( \chi_3 \) side chain dihedral angle distributions for the antiparallel and parallel \( \beta \)-sheet Q10 fibril polymorphs from MD simulations and UVRR spectra. Source: figure prepared with Python package Matplotlib and Adobe Illustrator.

UVRR spectra of NDQ10 and DQ10.

These results shed light on the structure of glutamine side chains in polyQ fibrils. Both computation and experiment report \( \chi_3 \) side chain dihedral angles centered close to 0°. Figure 2.11 shows the orientation of interdigitated glutamine side chains in the fibril Models \( a \) and \( b \). One can observe that the side chains are approximately planar in orientation. This planarity allows the side chains from opposing stacked \( \beta \)-sheets to tightly interdigitate, forming the hallmark steric zipper. Interestingly, this conformation also enables the primary amide groups on the side chains to accept and donate H-bonds between the side chains on adjacent \( \beta \)-strands, forming an organized network of H-bonds. This is the first analysis of polyQ fibril side chain structure that predicts the side chain zipper interactions that have been abundantly reported in the literature.

2.4 Conclusions and Direction

In this study, we performed a detailed structural analysis of fibrils formed from two conformational states of Q10 using MD paired with UVRR spectroscopy. Analy-
sis and comparison of Ψ dihedral angle distributions, as well as analysis of structures from the MD simulations led us to conclude that Q10 fibrils are composed of extended β-strands, rather than β-hairpin structures proposed in the literature. The predominant type of structure observed in Q10 fibrils is an ordered antiparallel β-sheet. Computational free energy calculations showed that antiparallel β-sheets are significantly lower in free energy than parallel β-sheets. This is an important finding that helps explain why polyQ fibrils prefer antiparallel configurations, in contrast to most types of amyloid fibrils. We performed a H-bonding analysis of antiparallel and parallel β-sheet fibril models that indicates that the favorability of antiparallel structures arises from H-bonding, particularly the formation of a greater number of favorable backbone-backbone H-bonds by antiparallel β-sheets. Finally, analysis of

Figure 2.11: Illustration of planar side chain H-bonding network. Side chains from top and bottom β-sheets form tightly packed cores. Source: figure prepared with Matplotlib and Adobe Illustrator.
distributions of the $\chi_3$ side chain dihedral angle definitely showed that Q10 side chains prefer to form planar conformations that can tightly interdigitate, which gives rise to the steric zipper interactions that stabilize polyQ fibrils.

This investigation of Q10 fibril structure has shown that, for short polyQ peptides, extended antiparallel $\beta$-sheets are the predominant structural motif in polyQ fibrils. The impact of this discovery lies in the understanding of the aggregation mechanism for Q10, and we have shown that the aggregation mechanism must lead to the formation of extended antiparallel $\beta$-sheets rather than $\beta$-hairpin fibrils. A remaining question was the monomeric phenomenon that lead to the aggregation and fibril formation of Q10. Xiong et al.$^{25}$ had shown that Q10 can adopt two distinct conformational states as an aqueous monomer. The next step in this dissertation was to develop a molecular picture of these monomeric states for Q10 to better elucidate beginning steps of the aggregation process depicted in Figure 1.3.
Chapter 3. CALCULATION AND EXPLORATION OF THE MONOMERIC
CONFORMATIONAL ENERGY LANDSCAPE OF Q10 PEPTIDES

This section is partially adapted and reproduced (with permission) from the following manuscript: Punihaole, D.; Jakubek, R.S.; Workman, R.J.; Marbella, L.E.; Campbell, P.; Madura, J.D.; Asher, S.A. The Journal of Physical Chemistry B. 2017, 121(24), 5953-5967.

3.1 Review of Aqueous PolyQ Monomeric Structure

All nine known polyQ diseases are caused by elongated CAG codon repeats that encode expanded polyQ tracts in proteins.\textsuperscript{10,11} In these diseases, expanded polyQ protein tracts are the only shared commonality that links the disease pathologies.\textsuperscript{13,95} The residue sequences that flank the polyQ repeat sections have been reported to influence the kinetics of the aggregation mechanism,\textsuperscript{27,38} but the flanking sequences vary by polyQ disease yet all display the same molecular pathology. This indicates that the properties of the polyQ repeat expansions alone are responsible for causing native proteins to misfold and aggregate into amyloid-like fibrils. Because of this, investigating the structural tendencies of polyQ peptides can aid in understanding how these peptides disrupt protein structure and subsequently aggregate. It has been shown that polyQ aggregates prepared in vitro from model polyQ peptides composed of only glutamine residues form fibrils that are remarkably similar to those formed in the brains of polyQ disease patients.\textsuperscript{8,20} In particular, in vitro fibrils composed of model Q\textsubscript{N} peptides of varying repeat length share properties such as filamentous morphologies, binding amyloid markers such as thioflavin-T, and characteristic $\beta$-sheet-rich fibril structures.\textsuperscript{18,96}

Many experimental studies of soluble polyQ peptide structure report that these peptides adopt structurally disordered conformations.\textsuperscript{18,75,77,96–98} However alternative conclusions have been drawn, one such study asserts that polyQ peptides contain
small sequences of folded structures that are thought to possess cytotoxic qualities. In addition, numerous computational investigations of these peptides have reported conflicting structural propensities. For example, Singh et al. report that polyQ peptides resist collapse and exist in primarily extended conformations, while studies by Wang et al. and Vitalis et al. observe collapsed disordered structures that are poorly hydrated. One thing that all studies echo is the lack of regular secondary structures in monomeric polyQ, with only transient portions of β-helices, β-sheets, or β-turns.

An important distinction to be made here is that, although polyQ peptides are thought to have disordered structures, most studies do not report behavior indicative of a true random coil polymer. A number of studies find that end-to-end distances of polyQ peptides differ significantly from random coil models in that they favor collapsed conformations, where a random coil has no preference. One experimental study used CD and NMR to show that short polyQ sequences have a high propensity to form polyproline II-like (PPII-like) structures.

A variety of key questions about polyQ monomeric structure elude the field. For example, and of critical importance in this dissertation, the behavior of these peptides that leads to fibril nucleation and growth is not understood. Chen et al. have proposed that fibril nucleation is triggered by an unfavorable conversion from random coil structure to β-sheet. Alternatively, work by Vitalis et al. posits that oligomers of these peptides form disordered, globular aggregates that undergo a structural conversion into ordered β-sheet fibrils.

Gaining a detailed understanding of the conformational ensemble that governs the monomeric structural tendencies of polyQ peptides would provide insight that could illuminate many current discrepancies in the field. PolyQ monomeric structure is key to characterizing the initial steps of the aggregation process (top row in Figure 1.3), as monomeric conformations adopted by polyQ peptides will determine the association.
events that induce aggregation (middle and bottom rows of Figure 1.3). Many of
the contradictions and disagreements concerning monomeric polyQ structure arise
from the limitations of the standard biophysical analysis techniques used to study
these systems. A validated pairing of multiple analysis techniques, optimally both
experimental and computational, is desired.

3.2 MD Simulations and UVRR Spectroscopy of Aqueous Q10 Monomers

In this dissertation, we characterized the structural ensemble of monomer $D_2Q_{10}K_2$
(Q10) peptides using an approach that synergistically utilizes MD simulations and
UVRR spectroscopy. This work built upon initial observations made by Xiong et al.$^{25}$
We performed an in silico quantification and exploration of the conformational free
energy landscape for Q10 using metadynamics MD simulations. Structures populating
local and global minima on this energy landscape were extracted from the simulation
data for further study. As in Chapter 2, backbone $\Psi$ and side chain $\chi_3$ dihedral angles
were calculated for these structures, and the resultant distributions were compared
to analogous distributions obtained from UVRR experiments of Q10 monomers. A
detailed H-bonding characterization of the different monomeric structural states was
performed using the simulation data.

Metadynamics MD simulations$^{43,66}$ were employed to calculate and characterize
the Q10 conformational free energy landscape. The initial molecular structure for
Q10 simulated here was a fully extended peptide with $(\Phi, \Psi)$ backbone dihedral an-
gles of $180^\circ$. This peptide was solvated with 6681 water molecules to create a periodic
simulation box with dimensions of $60 \times 60 \times 60 \text{ Å}^3$. This initial system was energy
minimized for 10000 steps using the conjugate gradient minimization scheme, fol-
lowed by 50 ps of equilibration under constant temperature and pressure of 300 K and
1.013 25 bar. After equilibration, metadynamics MD simulations$^{43,66}$ were performed
for a total of 1 µs. Data from these simulations was used to construct the free energy
landscape shown in Figure 3.2.
The TIP3P water model was used to model water in all simulations, along with the particle mesh Ewald algorithm for full-system electrostatics with a grid spacing of 1.0 Å. The Verlet velocity algorithm was used to integrate the equations of motion with a time step of 2 fs. All simulations were performed under constant pressure and temperature (NPT) with a Langevin thermostat and piston utilized to maintain a temperature of 300 K and pressure of 1.01325 bar respectively. A pair interaction cutoff of 12.0 Å was used with a nonbonded switching distance of 10.0 Å. All simulation analysis was performed with the VMD software package.

To specify the configurational phase space for the energy landscape, we used three RMSD collective variables that reported the conformation of the Q10 α-carbons relative to three idealized structures: α-helix, α-hairpin, and PPII-helix. The α-helix and PPII structures were created using MOE software, and the β-hairpin was taken from previous Q10 MD simulations. A maximum RMSD value of 12.0 Å was used and this upper limit was maintained with a harmonic potential of the form \( \frac{1}{2}k(x-x_0)^2 \) with a force constant, \( k \), of 1.0 kcal mol\(^{-1}\). This approach for calculating the conformational energy landscape was employed by Gaborek et al. previously.

In the metadynamics simulation, Gaussian energy functions of height \( \sim 1.0 \text{ kcal mol}^{-1} \) and width 3.0 Å were added to the energy landscape every 500 steps (see Equation 1.10). Low energy structures were extracted from the metadynamics simulation based on their RMSD values and the history of deposited Gaussians. More Gaussians deposited at a certain conformation correspond to a lower energy structure. These low energy structures were simulated using standard MD without metadynamics for 10 ns to collect equilibrium statistics. The H-bonding present in these trajectories was then analyzed using the H-bonding module in VMD. A heavy atom distance of <3.5 Å and H-bond angles of 180 ± 30° were used to determine the presence of a H-bond. \( \Phi \) and \( \Psi \) backbone dihedral angles were gathered from the simulation trajectory with a Tcl script in VMD. Free energy output from the metadynamics
calculation and associated RMSD collective variable values were processed with two Python scripts, and the R statistical computing package was used to plot the energy landscape.\textsuperscript{94}

Constrained MD simulations were used to evaluate the potential energy of Q10 in PPII and 2.5\textsubscript{1}-helix conformations, constructed with (\(\Phi, \Psi\)) angles of \((-75^\circ, 150^\circ)\) and \((-130^\circ, 177^\circ)\) respectively. In these simulations, constraints were imposed on the \(\Phi\) and \(\Psi\) dihedrals angles with NAMD.\textsuperscript{43} This caused the 2.5\textsubscript{1}-helix and PPII peptides to keep their secondary structures throughout the simulations. 0.25 kcal mol\(^{-1}\) \AA\(^{-1}\) force constants were placed on the dihedral angles with the Collective Variables module in NAMD.\textsuperscript{43} In the constrained simulations, solvated Q10 systems were energy minimized for 10 000 steps using the conjugate gradient method,\textsuperscript{43} followed by 1 ns of equilibration. Data collection MD runs of 200 ns were then carried out. Potential energies were obtained with the NAMD Energy module.\textsuperscript{43} Stretching, bending, torsional, van der Waals, and Coulombic energies were computed for every frame of the coordinate trajectory files and averaged using a Python script. H-bonding analysis of the trajectories was performed with VMD’s Hydrogen Bond tool.

3.3 Characterization of the Monomeric Q10 Conformational Ensemble

UVRR spectra of the NDQ10 and DQ10 peptide samples shows that DQ10 populates a structural state that is PPII-rich. This state also has a \(\Psi\) angle distribution with a peak at 170\(^\circ\), which is not a common region for \(\Psi\) angles to occupy.\textsuperscript{103} Xiong et al. observed this peak previously and structure resulting from this has been deemed a 2.5\textsubscript{1} helix.\textsuperscript{25} Both PPII and 2.5\textsubscript{1}-helix represent extended structures, so the DQ10 state can be characterized as an extended state rich in PPII-helix and featuring a fraction of 2.5\textsubscript{1}-helix. NDQ10 on the other hand has a clean \(\Psi\) angle distribution with a large peak centered around 140\(^\circ\) which corresponds to \(\beta\)-strand type structure. The \(\Psi\) angle distributions from the UVRR experiments are shown in Panels a and c of Figure 3.1.
Figure 3.1: Distributions for peptide Ψ and χ₃ dihedral angles from MD simulations and UVRR experiments. Panels a and c show UVRR-obtained Ψ angles for DQ10 and NDQ10 peptide samples, and panels b and d show MD-calculated Ψ distributions for states B (collapse β-strand) and A (PPII). The right side shows χ₃ distributions organized the same as the right side. Source: figure prepared with Python package Matplotlib and Adobe Illustrator.

Data from the metadynamics MD simulations was used to gain insight into the relative energies of the Q10 structural, as well as the activation barrier between the DQ10 and NDQ10 monomer states. Figure 3.2 shows the conformational Gibbs free energy landscape of Q10 as a function of three collective variables, defined as RMSD comparisons to three idealized structures: α-helix, β-hairpin, and PPII. These were chosen because they are common structural motifs that enable one to distinguish
clearly between possible Q10 conformations. Additionally, a β-hairpin structure was used as this structure has been implicated as a possible aggregation nucleus.30,74

We observe that the conformational free energy landscape of Q10 is rugged and frustrated, as it contains a many relatively shallow local minima. However, four deep free energy minima (dark purple areas on Figure 3.2) were identified. The energy minima contain different structural states (A-D) that are defined both by their backbone dihedral angles (Figure 3.3) and their H-bonding properties (Figure 3.4).

The structures populating State A possess backbones with average (Φ, Ψ) angles of (−90°, 140°), in the region of a β-strand-like conformation (Figure 3.3a). These conformations are referred to as “collapsed” β-strands, instead of β-hairpins, because they lack the backbone amide H-bonding patterns that define canonical β-hairpins. As evidenced in Figure 3.4, these β-strand structures form, on average, more side chain↔backbone H-bonds than the other monomeric states. These side chain↔backbone H-bonds, together with the attractive charge-charge interactions between the aspartate and lysine residues, cause the β-strands to adopt compact (“collapsed”) conformations.

The (Φ, Ψ) angle distribution for State B has a strong population at about (−80°, 150°), which is indicative of PPII secondary structure (Figure 3.3b). A minority of Ramachandran angles, located around (−50°, −57°), indicate that State B also contains β-turn-like conformations. Viewed together, these two populations of (Φ, Ψ) angles imply that State B is made up of structures containing short PPII helices broken up by turns or bends. Our H-bonding analysis of this state (Figure 3.4) indicates that, compared to State A, both the backbone and side chains of State B are preferentially hydrated by peptide↔water H-bonds.

The structures that reside in State C have backbone (Φ, Ψ) angles centered at (−135°, 135°), characteristic of β-strand structures (Figure 3.3c). A minor population adopts type I’ β-turn angles (−40°, −85°). In contrast to State A’s collapsed
$\beta$-strand structures, $\beta$-hairpin structures found in State C form more intrapeptide backbone↔backbone H-bonds(Figure 3.4c). These backbone H-bonds are indicative of canonical $\beta$-hairpin secondary structures, as defined by Milner et al.$^{104}$ Additionally, the side chains of these $\beta$-hairpins are more solvated by H-bonds to water than those of the collapsed $\beta$-strand structures populating State A (Figure 3.4a).

The Q10 conformations located in State D are $\alpha$-helical conformations. As shown
Figure 3.3: $(\Phi, \Psi)$ dihedral angle distributions for the (a) collapsed $\beta$-strand structures from State A, (b) PPII structures from State B, (c) $\beta$-hairpin structures from State C, and (d) $\alpha$-helix structures from State D. Panel (e) shows $(\Phi, \Psi)$ angles of terminal aspartate and lysine residues from State B while (f) shows $(\Phi, \Psi)$ angles from all glutamine residues from State B. Source: figure prepared with Python package Matplotlib and Adobe Illustrator.
in Figure 3.3e, the $(\Phi, \Psi)$ angles are distributed uniformly $(-50^\circ, -65^\circ)$. This structural state possesses a large amount of backbone↔backbone H-bonding (Figure 3.4c), as expected for $\alpha$-helix secondary structure, while the side chains remain well solvated.

The collapsed $\beta$-strand (State A) occupies the global minimum in the conformational free energy landscape, while the PPII-rich (State B), $\beta$-hairpin (State C), and $\alpha$-helix (State D) states occur at 0.5, 0.66, and 1.0 kcal mol$^{-1}$ higher in energy, respectively. A 3 kcal mol$^{-1}$ to 6 kcal mol$^{-1}$ energy barrier separates states A from state B. Both states C and D are isolated from all other states by a 6 kcal mol$^{-1}$ to 15 kcal mol$^{-1}$ energy barrier.

Our analysis of the backbone $\Psi$ and $\Phi$ data for States A-D (Figure 3.3a-d) concludes that State A (collapsed $\beta$-strand) and State B (PPII-like) structures correspond to the UVRR measured NDQ10 and DQ10 monomers, respectively. To validate the simulation observations, the UVRR measured $\Psi$ and $\chi_3$ angle distributions of DQ10 and NDQ10 were compared to the distributions calculated from the Q10 structures populating the A and B structural states (Figure 3.1). Agreement between the experimental and simulation $\Psi$ angle distributions is excellent. However, the $\chi_3$ angle distributions for the side chains taken from the MD simulation data agrees poorly with the UVRR distributions.

Figure 3.1a-d compares $\Psi$ distributions of States A and B from the metadynamics simulations to those of experimental DQ10 and NDQ10. The corresponding $\Psi$ angle distributions are in good qualitative and quantitative agreement. The only significant discrepancy is that the distribution for State B (Figure 3.1b) lacks a peak at $\sim 175^\circ$ that corresponds to experimentally observed 2.5$_1$-helix conformations. This incongruence is discussed in detail later.

The $\chi_3$ distributions for states A and B calculated from the metadynamics trajectories are bimodal (Figure 3.1f,h). Both distributions have large peaks around $\sim -70^\circ$ and $\sim 110^\circ$. These $\chi_3$ angles do not agree with those measured by UVRR. The side
chain rotamer database by Dunbrack et al.\textsuperscript{105} reports that glutamine residues normally possess $\chi_3$ angles distributed between $\sim$−90° and 90°, inconsistent with the distributions shown in Figure 3.1f,h.

We suspect that the disparity between our MD simulated $\chi_3$ angle distributions and those calculated from the UVRR data stems from the lack of parametrization for non-rotameric side chain dihedrals in the CHARMM36 force field.\textsuperscript{50} Best et al.\textsuperscript{106} document how the $\chi_1$ and $\chi_2$ dihedral angles of various side chains were parameterized in CHARMM36 using experimental data. However, to our knowledge, non-rotameric side chain dihedral angles, such as the $\chi_3$ angle of glutamine, have not been similarly optimized in CHARMM36.\textsuperscript{106}

We compared the backbone and side chain H-bonding populations present in the metadynamics structural states (Figure 3.4).

Figure 3.4: Depicted here are H-bonding populations for the four structural minima identified in Figure 3.3. Only amide-amide or amide-water H-bonds were considered. Source: figure prepared with Python package Matplotlib and Adobe Illustrator.

Our computational results mirror the implications of our experimental findings.

43
The State B structures feature primarily peptide↔water H-bonding, while the State A structures show significant intrapeptide H-bonding.

Our metadynamics data confirm the presence of a significant free energy activation barrier between the PPII (DQ10) and the collapsed β-strand (NDQ10) conformational states. We calculate an energy barrier of 3-6 $k_B T$ separating states A (β-strand) and B (PPII). The energy barrier region between states A and B also contains $\geq 8$ local energy minima, with relative energy well depths of $\sim 1.0$ kcal mol$^{-1}$ (Figure 3.5). These local minima represent metastable structural states along the energy barrier. Even if Q10 can surmount the large energy barrier, the kinetics of transition will be impeded by these energy wells to prevent PPII-like to β-strand structural interconversion.

The UVRR spectra of DQ10 surprisingly features a significant population of the $2.5_1$-helix conformation previously proposed by Krimm and co-workers$^{67}$. We used classical MD simulations to more deeply examine the factors that stabilize $2.5_1$-helix conformations in Q10. We simulated three different Q10 peptide structures. One structure has its backbone dihedral angles constrained to the canonical $2.5_1$-helix conformation, the second structure has a PPII conformation, and the third structure has alternating PPII and $2.5_1$-helix backbone dihedral angles. We identified the overall lowest energy structure of each of the constrained simulations. We then calculated, for each Q10 structure, the energy contributions of bond stretches, bond angle bending, dihedral angle rotation, electrostatics, and van der Waals to the overall potential energy of the system (Table 3.1). Neglecting the influence of solvating waters, our simulations indicate that the potential energy of the $2.5_1$-helix is $\sim 50$ kcal mol$^{-1}$ lower than that of the PPII structure, mainly due to intrapeptide electrostatics.

To determine the electrostatic potential energy contributions of the charged terminal residues, we compared the electrostatic energy terms of just the glutamine residues of Q10 for the $2.5_1$- helix and PPII structures. We find that for glutamine residues the electrostatic potential energy difference between the $2.5_1$- helix and PPII
Figure 3.5: Shown is the Q10 monomeric conformational energy landscape with the energy barrier region between states A and B highlighted. Local energy minima on the barrier are shown as an orange energy surface, and example local minimum structures are shown as (a) and (b). Source: figure prepared with VMD, R, and Adobe Illustrator.

Table 3.1: Contributions (in kcal mol$^{-1}$) of various types of potential energies to the total energy of entirely PPII monomeric Q10, partially PPII/2.5$_1$-helix Q10, and entirely 2.5$_1$-helix structure. Columns labeled “all” contain potential energies taking glutamine and terminal residues into account, while columns labeled “polyQ” contain potential energies calculated just for glutamine residues.
structures is only $\sim 10 \text{kcal mol}^{-1}$. Thus, the charged terminal residues contribute most heavily to the electrostatic potential energy difference between the PPII and $2.5_1$-helix structures.

We also compared the ($\Phi$, $\Psi$) angle distributions of the glutamine residues and the charged flanking residues for State B structures. As shown in Figure 3.3e, the terminal aspartic acid and lysine residues preferentially adopt ($\Phi$, $\Psi$) angles centered roughly at $(-115^\circ, -160^\circ)$, closer to the ($\Phi$, $\Psi$) angles of the canonical $2.5_1$-helix conformation. In contrast, the glutamine residues (Figure 3.3c) adopt ($\Phi$, $\Psi$) angles of $(-80^\circ, -150^\circ)$ and $(-50^\circ, -57^\circ)$, indicative of PPII and turn-like conformations, respectively. This suggests that the $2.5_1$-helix-like conformations are selectively localized to the peptide bonds of the charged terminal residues, while the glutamine residues occur in a predominately PPII conformation. Our collaborators calculate that a single DQ10 peptide has 2 residues to 3 residues in a $2.5_1$-helix conformation, most likely occurring on the terminal residues. However, our MD simulations indicate that the $2.5_1$-helix-like $\Psi$ angles present in State B are centered around $-160^\circ$, closer to the values for PPII structures than the value of $-175^\circ$ measured by the UVRR.

We attribute the discrepancy in the $\Psi$ angle values for the $2.5_1$-helix-like peptide bonds in the simulated State B structures to an inadequacy of the CHARMM36 force field. In the case of CHARMM36, parameters for backbone dihedral angles are optimized from globular protein X-ray crystal structures and QM data for the dialanine peptide.\textsuperscript{55} The $2.5_1$-helix is an uncommon structure that forms in the presence of adjacent charged amino acids and is unlikely to be well represented in the X-ray structures used to parametrize CHARMM36. This conclusion is supported by Liqi Feng, who showed that metadynamics of poly(L-lysine), using the CHARMM36 force field, failed to show a $\Psi$ angle distribution at $\sim 170^\circ$, indicative of $2.5_1$-helix conformations.\textsuperscript{107}
3.4 Discussion and Conclusions

The solution-state structures of polyQ-rich peptides and proteins have been studied in detail.\textsuperscript{19,75–77,100,108} Many of the important computational studies were conducted by Pappus group.\textsuperscript{19,76,77} For example, Wang et al.\textsuperscript{76} characterized the conformational ensemble of N-acetyl-Q5-N'-methylamide (Q5) and N-acetyl- Q15-N'-methylamide (Q15) that are somewhat similar to our Q10 peptide.

From the MD simulations performed by Wang et al.,\textsuperscript{76} they conclude that the conformational energy landscapes of Q5 and Q15 are frustrated, since they observe numerous metastable and glassy states. Their simulations find essentially no global energy minimum structures. They show that Q5 and Q15 are structurally disordered in aqueous solution, with only transient stretches of regular secondary structure elements, such as PPII-helices, $\beta$-strands, and $\alpha$-helices occurring. They also show that both Q5 and Q15 peptide bonds have a strong propensity to adopt PPII-like and $\alpha$-helix-like ($\Phi$, $\Psi$) angles.

Wang et al.’s simulations also suggest that structurally disordered, monomeric polyQ peptides form a significant number of interamide H-bonds, with side chain$\leftrightarrow$backbone H-bonding being the most prevalent.\textsuperscript{76} This is supported by the NMR measurements by Darnell et al.\textsuperscript{109} that find that in PPII-rich polyQ peptides the glutamine side chains adopt “folded-over” conformations that enable H-bonding to the backbone amides. To our knowledge, this is the only experimental evidence of significant side chain$\leftrightarrow$backbone H-bonding in PPII-rich polyQ peptides.

The simulations of Wang et al.\textsuperscript{76} lead them to hypothesize that the structural disorder of polyQ peptides stems from the many different possible combinations of intramolecular and intermolecular-amide H-bonds that can form between side chain and backbone amides. This explains why a $\beta$-sheet-rich fibril nucleus\textsuperscript{13,16} is expected to be energetically unfavorable. According to their model, the multiple possible combinations of intrapeptide H-bonds promote disorder in the peptide backbone, disrupting
the formation of backbone↔backbone interactions that enable secondary structures such as α-helices and β-sheets.

Some aspects of our simulation results are consistent with those of Wang and coworkers. We observe that the conformational energy landscape of Q10 contains numerous shallow energy minima. Similarly, we observe that Q10 peptides can engage in a large variety of intramolecular side chain↔side chain, backbone↔backbone, and side chain↔backbone H-bonding interactions. However, there are many aspects of our results that are inconsistent with their results. For example, our experimental and metadynamics simulation data reveal that Q10 can adopt stable and well-defined structural states, with deep energy wells, such as those shown in Figure 3.2. Interestingly, we also find that the side chains and backbone amides of the structurally “disordered” PPII-rich state of Q10 are predominately H-bonded to water, which disagrees with the Wang et al. results.

These disagreements raise two important questions. First, why does Wang et al. not observe any well-defined structural states for small polyQ peptides, whereas we do? And second, why do the simulations of Wang et al. indicate that there are significant numbers of interamide H-bonds (particularly between the backbone and side chain amides) in structurally disordered Q5 and Q15 peptides, while our simulations do not find this for Q10?

To answer the first question, we note that the conformational energy landscape of polyQ peptides is frustrated, consisting of many local energy wells. Classical MD simulations generally do not efficiently sample different structural states in these situations. Thus, one possibility is that the simulated peptides in the Wang et al. study are trapped in local energy minima. Our metadynamics calculations enabled us to sample robustly a greater ensemble of structures than does traditional MD simulations. Thus, we characterized the entire conformational landscape and discovered global and local minimum energy conformations.
We are also aware that the conformational behavior of model polyQ peptides in solution depends on both the number of glutamine repeats and the choice of non-glutamine flanking groups. For example, the use of N-acetyl and methylamide flanking groups in Pappu and co-workers’ simulations\textsuperscript{19,76,77} are better models for polyQ tracts in proteins. In addition, fluorescence correlation spectroscopy data\textsuperscript{75} by the Pappu group and dynamic light scattering data by the Murphy group\textsuperscript{17} indicate that structurally disordered long polyQ peptides adopt relatively compact structures in aqueous solution. These compact structures presumably derive from the fact that interamide H-bonding interactions become more prevalent and important in polyQ peptides as the glutamine repeat length increases. For example, Walters and Murphy\textsuperscript{17} have shown that phosphate buffer is a good solvent for $Q_8$ and $Q_{12}$, a mediocre solvent for $Q_{16}$, and a poor solvent for larger peptides such as $Q_{20}$. Thus, the model proposed by Pappu and co-workers in the Wang et al.\textsuperscript{76} study may be valid in the limit of polyQ peptides with $\geq 20$ glutamine repeats.

The answer to the second question may lie in the differences of the force field and water model used by our study compared to the Wang study. As discussed by Wang et al.,\textsuperscript{76} the H-bonding interactions observed in a simulation will vary depending on the force field utilized. Our use of the TIP3P water model and the CHARMM36 modern force field results in different H-bonding interactions compared to that of the MD simulations of Wang et al.\textsuperscript{76} The H-bonding interactions predicted by our metadynamics are in excellent agreement with our UVRR data, which show that our choice of force field appropriately models the H-bonding interactions in Q10.

The congruence of both our simulation and experimental data gives us confidence in the relative accuracy and robustness of the CHARMM36\textsuperscript{50} force field used in this study. Despite this, we do not dispute the validity of the Pappu et al.’s model to describe, in general, the underlying physical principles that govern the structural disorder of larger polyQ peptides or the energetic unfavorability of the coil to $\beta$-sheet
transition. However, we do believe that the importance of side chain↔side chain and side chain↔backbone interactions in describing the structural disorder of PPII-rich polyQ peptides may be overstated, at least in the context of small peptide systems, such as Q5, Q10, and Q15.

In this chapter, we have shown that the monomeric Q10 structural ensemble, which corresponds to the top row of the aggregation mechanism shown in Figure 1.3, is made up of two conformational states: a PPII-rich structure and a collapsed β-strand. These findings are at odds with the right pathway in Figure 1.3 due to the lack of an accessible β-hairpin structure, and indicate that the leftmost and middle pathways better describe the Q10 aggregation mechanism.
Chapter 4. ANALYSIS OF RELATIVE HYDROGEN BONDING STRENGTHS FOR Q10 PEPTIDES AND FIBRILS

This section is partially adapted and reproduced (with permission) from the following manuscript: Punihaole, D.; Jakubek, R.S.; Asher, S.A. *The Journal of Physical Chemistry Letters.* 2018, 9(8), 1944-1950.

4.1 Background on the Role of H-bonding in PolyQ Systems

PolyQ peptides contain both primary amides from their glutamine (Q) side chains and secondary amides from their backbone peptide bonds. Despite the hydrophilic nature of the glutamine side chain, experimental studies indicate that polyQ peptides with pathologically relevant repeat lengths adopt structurally disordered collapsed conformations, which suggests that water is acting as a poor solvent.\textsuperscript{17,18,96–98,110} These findings are also supported by computational studies,\textsuperscript{19,76,77,101} which suggest that polyQ peptides are largely disordered due to the multiplicity of different H-bonding interactions possible between side chain and backbone amides. Other computational studies suggest that interamide H-bonds between side chains contribute most significantly to the structural stability of polyQ amyloid-like fibrils.\textsuperscript{111}

These and other studies\textsuperscript{11,32} emphasize the role that glutamine side chain H-bonding interactions play in dictating the solution-state conformational behavior and the strong aggregation propensities of polyQ peptides. Surprisingly, no experimental studies have quantified the relative energetic favorability of side chain versus backbone amide H-bonding interactions in polyQ peptides. Thus, developing new tools that can quantify the relative energies of different side chain and backbone H-bonding interactions is important to formulating a more complete, molecular-level understanding of polyQ fibril formation mechanisms. An important unresolved question to answer in the field is whether the peptidepeptide H-bonding of polyQ peptides is energetically stronger than peptidewater interactions.
4.2 Calculating the Interaction Enthalpy of Q10 H-bonding with MD and UVRR

Our collaborators developed a method with which to estimate the change in interaction enthalpy that occurs when a secondary amide forms different types of H-bonds relative to the amide in vacuum.\(^{35}\) The frequency of peptide amide vibrational modes, particularly the secondary amide I (AmI\(^S\)) band arising from carbonyl stretching, show sensitivity to the H-bonding environment of peptide amides. Wang et al.\(^{112}\) previously showed that the AmI\(^S\) frequency is linearly dependent on the solvent acceptor number, which is a measure of the strength of the H-bonding and van der Waals interactions of the solvent with the solute. Using this linear correlation, Wang et al. showed that the AmI\(^S\) frequency can be used to estimate the change in interaction enthalpy (\(\Delta H_{\text{int}}\)) that occurs when a secondary amide is placed into a solvent environment.

Using this method, \(\Delta H_{\text{int}}\) values were calculated from UVRR spectra for various types of side chain H-bonds in Q10 solutions. Several types of H-bonds were calculated, including side chain↔water H-bonds for the PPII monomeric state (DQ10), side chain↔backbone and side chain↔water H-bonds for the collapsed \(\beta\)-strand monomeric state (NDQ10), and side chain↔side chain H-bonds for Q10 fibrils prepared from both the PPII and collapsed \(\beta\)-strand states.

We used MD simulations to calculate analogous \(\Delta H_{\text{int}}\) values for the same types of H-bonds. In Chapters 2 and 3, MD simulations of Q10 \(\beta\)-sheet fibrils and monomeric structures were reported, respectively. In this work, molecular structures from each of these previous studies were simulated with classical MD, and data from these trajectories was used to obtain \(\Delta H_{\text{int}}\) measurements for the H-bonding environments. From Chapter 2, the antiparallel \(\beta\)-sheet model a was used to calculate H-bonding enthalpies to compare with the UVRR fibril measurements, and from Figure 3.2 in Chapter 3, molecular structures for states A and B were used to compare with their
corresponding experimental structural states NDQ10 and DQ10. The three structures simulated here are shown in Figure 4.1.

Figure 4.1: (a) Collapsed $\beta$-hairpin Q10 monomer (state A from 3), (b) PPII-rich Q10 monomer (state B from 3), (c) Octameric antiparallel extended $\beta$-sheet fibril (model a from 1). Source: figure prepared with VMD$^{83}$

Coordinates for Structures $a$ and $b$ from Figure 4.1 were taken from the work performed in Chapter 3, and Structure $c$ was taken from the Q10 fibril study in
Chapter 2. Monomeric Structures $a$ and $b$ were previously solvated in $60 \times 60 \times 60 \, \text{Å}^3$ periodic water boxes composed of 6681 water molecules. The antiparallel Q10 fibril structure (Structure $c$) was solvated in a periodic $70 \times 50 \times 50 \, \text{Å}^3$ water box made up of 5087 water molecules.

Each system was simulated for 5 ns under constant temperature and pressure using NAMD.$^{43}$ The CHARMM36 force field,$^{50}$ which has proven to be accurate for Q10 systems,$^{33,34}$ was used to calculate potential energies and forces. The TIP3P$^90$ water model was used to model water in all simulations, along with the particle mesh Ewald algorithm$^{91}$ for full-system electrostatics with a grid spacing of 1.0 Å. The Verlet velocity algorithm$^{45}$ was used to integrate the equations of motion with a time step of 2 fs. All simulations were performed under constant pressure and temperature (NPT) with a Langevin thermostat and piston utilized to maintain a temperature of 300 K and pressure of 1.013 25 bar respectively. A pair interaction cutoff of 12.0 Å was used with a nonbonded switching distance of 10.0 Å. All H-bonding $\Delta H_{\text{int}}$ analyses were performed with the Tcl scripting interface in VMD.$^{83}$

We defined $\Delta H_{\text{int}}$ as the sum of the Lennard-Jones ($\Delta E_{\text{int}}^{\text{LJ}}$) and Coulombic, or electrostatic, potential energy ($\Delta E_{\text{int}}^{\text{elec}}$) terms, where the $\Delta$ signifies the energy difference between an amide interaction at close distance and infinite separation. Due to the negligible change in volume throughout the MD simulations, $\Delta H_{\text{int}}$ can be accurately approximated as the change in internal energy of interaction ($\Delta E_{\text{int}}$). The mathematical relationship is shown below.

$$\Delta H_{\text{int}} \approx \Delta E_{\text{int}} = \Delta E_{\text{int}}^{\text{elec}} + \Delta E_{\text{int}}^{\text{LJ}}$$  \hspace{1cm} (4.1)

When calculating $\Delta H_{\text{int}}$, we define interacting groups as those with heavy atoms at distances of less than or equal to 5 Å. Thus, our calculated $\Delta H_{\text{int}}$ is not limited to strong H-bonding interactions, which generally occur for heavy atom dis-
stances less than 3 Å. For the Q10 antiparallel β-sheet conformation (Figure 4.1c), the six innermost buried side chain and backbone amide groups were used to calculate the $\Delta H_{\text{int}}$ because they best model the interior of the fibril core.

### 4.3 Evaluating the Comparative Strengths of Q10 Side Chain H-bonds

Table 4.1 contains the $\Delta H_{\text{int}}$ calculated for all the systems with both MD and UVRR experiments. From experiment, $\Delta H_{\text{int}}$ values were calculated for solutions of aqueous glutamine whose side chains are H-bonding with aqueous solvent. They calculated a $\Delta H_{\text{int}}$ value of $-4.3 \text{ kcal mol}^{-1}$ for these glutamine water H-bonding interactions. This served as a baseline of comparison for glutamine side chains H-bonding to water in Q10 peptides.

We calculated $\Delta H_{\text{int}}$ of side chain↔water H-bonding for the PPII structural state of monomeric Q10 (DQ10 or state B from Chapter 3) using both MD and UVRR. The agreement for these calculated values is very good; UVRR experiments calculate a value of $-4.2 \text{ kcal mol}^{-1}$ for PPII peptide↔water H-bonds while our MD experiments yield a $\Delta H_{\text{int}}$ of $-4.4 \text{ kcal mol}^{-1}$. This indicates that Q10 side chains in the PPII structural state are hydrated with the same energetic strength as the side chains of single glutamine residues in solution, because the side chain↔water $\Delta H_{\text{int}}$ values are similar between aqueous glutamine and PPII-rich Q10 monomers.

<table>
<thead>
<tr>
<th>System</th>
<th>Expt. $\Delta H_{\text{int}}$</th>
<th>MD $\Delta H_{\text{int}}$</th>
<th>H-bonding types</th>
</tr>
</thead>
<tbody>
<tr>
<td>glutamine</td>
<td>$-4.3$</td>
<td>$-4.4$</td>
<td>s.c.-w.</td>
</tr>
<tr>
<td>PPII monomer</td>
<td>$-4.2$</td>
<td>$-4.4$</td>
<td>s.c.-w.</td>
</tr>
<tr>
<td>β-strand monomer</td>
<td>$-5.8, -4.3$</td>
<td>$-4.7, -4.4, -4.6$</td>
<td>s.c.-s.c., s.c.-w., s.c.-b.b.</td>
</tr>
<tr>
<td>NDQ10 fibril</td>
<td>$-5.9$</td>
<td>$-6.2$</td>
<td>s.c.-s.c.</td>
</tr>
<tr>
<td>DQ10 fibril</td>
<td>$-5.4$</td>
<td>$-5.4$</td>
<td>s.c.-s.c.</td>
</tr>
</tbody>
</table>

Table 4.1: Side chain H-bonding $\Delta H_{\text{int}}$ values (kcal mol$^{-1}$) for different types of Q10 monomeric and fibril side chain interactions. MD simulations were not performed on monomeric glutamine, thus no MD data is presented for that system; also we had no way of obtaining an MD structure for fibrils prepared from PPII (DQ10) monomers. We judged the extended antiparallel β-strand fibril model to be analogous to fibrils prepared from NDQ10 peptides.

The H-bonding of aqueous Q10 peptides in β-strand conformations (NDQ10 or
state A from Chapter 3) was also analyzed. UVRR was used to calculate $\Delta H_{\text{int}}$ values for side chain↔backbone and side chain↔water H-bonds, and with MD we calculated the strength of these, as well as side chain↔side chain H-bonds. With UVRR, side chain↔backbone H-bonding in Q10 $\beta$-strands was estimated to have an enthalpic strength of $-5.8 \text{ kcal mol}^{-1}$, significantly higher than the $\Delta H_{\text{int}}$ for side chain↔water H-bonds, which was measured to be $-4.3 \text{ kcal mol}^{-1}$. MD-calculated $\Delta H_{\text{int}}$ values for $\beta$-strand Q10 H-bonding were $-4.7$, $-4.4$, and $-4.6 \text{ kcal mol}^{-1}$ for side chain↔backbone, side chain↔water, and side chain↔side chain H-bonds, respectively. UVRR and MD results agree well again for side chain↔water H-bonding, with values of $-4.3$ and $-4.4 \text{ kcal mol}^{-1}$. Both methods yield larger $\Delta H_{\text{int}}$ values for side chain↔backbone H-bonds than for side chain↔water, however UVRR estimates a significantly higher $\Delta H_{\text{int}}$ of $-5.8 \text{ kcal mol}^{-1}$ compared to $-4.7 \text{ kcal mol}^{-1}$ calculated from MD. However, the trend of side chain↔side chain H-bonds being stronger than those of side chain↔water is conserved in both data. MD also was able to calculate a $\Delta H_{\text{int}}$ value of $-4.6 \text{ kcal mol}^{-1}$ for side chain↔side chain H-bonds in $\beta$-strand Q10, indicating that these interactions are comparable in energy to side chain↔backbone H-bonds.

The enthalpic strength of Q10 fibril side chain H-bonding was also calculated here. In Chapter 2, it was concluded that Q10 fibrils formed from the $\beta$-strand (NDQ10) and PPII (DQ10) states were composed primarily of antiparallel extended $\beta$-sheets. Our collaborators calculated $\Delta H_{\text{int}}$ values from UVRR spectra for Q10 fibrils prepared from both of these monomeric structural states to observe any energetic differences. Computational $\Delta H_{\text{int}}$ values were obtained from MD simulations of the octameric antiparallel extended $\beta$-sheet fibril (model a from Chapter 2) we previously investigated. UVRR experiments yielded side chain↔side chain H-bond $\Delta H_{\text{int}}$ values of $-5.9 \text{ kcal mol}^{-1}$ for fibrils prepared from $\beta$-strand monomers and $-5.4 \text{ kcal mol}^{-1}$ for those prepared from PPII monomers. MD calculated side chain↔side chain $\Delta H_{\text{int}}$
values from the antiparallel β-sheet Q10 fibril were of strength $-6.2 \text{kcal mol}^{-1}$.

These data indicate that the fibrils prepared from β-sheet and PPII monomers are not entirely of an identical morphology; it seems that Q10 fibrils formed from β-strand monomers feature side chain↔side chain H-bonding that is more similar to the antiparallel β-sheet computational model than the fibrils formed from PPII monomers. This may indicate that Q10 fibrils formed from PPII monomers are composed of a more significant fraction of parallel β-sheets, which would form different side chain-H-bonding patterns.

4.4 Conclusions and Direction

Our computational and experimental results show that $\Delta H_{\text{int}}$ values of side chain↔side chain and side chain↔backbone interactions of Q10 monomeric and fibril structures are enthalpically more favorable than side chain↔water interactions. This quantitatively confirms the hypotheses of various studies\textsuperscript{32,76,77,113} that argue that intrapeptide H-bonding is the driving factor in the structural ensemble of monomer and aggregated polyQ peptides. Interestingly, our results also indicate that side chain↔side chain H-bonding is stronger in Q10 fibrils than backbonebackbone H-bonding. To our knowledge, this work is the first study that experimentally and computationally quantifies the energetic favorability of polyQ H-bonding interactions for polyQ monomers and fibrils. In future work, we move away from the Q10 model peptide in the direction of longer peptides with more physiological relevance. The same structural and H-bonding analyses performed in Chapters 2, 3, and 4 can be readily applied to these longer peptides.
Chapter 5. LENGTH DEPENDENT PROPERTIES IN THE STRUCTURE OF Q15 AND Q20 PEPTIDE MONOMERS AND FIBRILS

5.1 Overview of Length Dependent Properties in PolyQ Peptide Structure

The length of the polyQ tract in diseased proteins affects the severity of the disease.\textsuperscript{10,11} Patients with longer polyQ tracts have an earlier disease age-of-onset, and disease symptoms only occur if the proteins polyQ tract surpasses a critical length.\textsuperscript{10,15,114} For example, clinical presentation of Huntington’s disease is only observed in patients with a polyQ tract $\geq 36$ residues long in the huntingtin protein.\textsuperscript{11} Additionally, \textit{in vitro} studies of small polyQ peptides show that longer polyQ tracts increase aggregation rates; this increase in aggregation rate may be related to the disease age-of-onset.\textsuperscript{18,114}

Because of the dependence of aggregation rate and disease age-of-onset on the length of the polyQ tract, there is great interest in determining the structural differences between polyQ tracts of different lengths. Most experimental studies conclude that solution-state polyQ peptides are intrinsically disordered regardless of the polyQ tract length.\textsuperscript{18,96,110} However, some studies suggest the presence of small populations of secondary structure, which may play a role in aggregation and cytotoxicity.\textsuperscript{18,110,114} In addition, computational studies agree that polyQ peptides are intrinsically disordered regardless of repeat length.\textsuperscript{19,76,77,115}

In Chapter 3, we used metadynamics simulations paired with UVRR spectroscopy to investigate the solution-state structure of D$_2$Q$_{10}$K$_2$ (Q10) peptides. We found that Q10 can exist in a PPII-rich structure (referred to as DQ10) or a collapsed $\beta$-strand conformation (NDQ10). There is a large activation barrier preventing interconversion of these two structural states, and, interestingly, DQ10 was found to be more resistant to fibrillization than NDQ10.
5.2 Investigation of Q15 and Q20 monomeric structural properties with MD and UVRR

Here, we utilized metadynamics MD simulations of monomeric D2Q15K2 (Q15) and D2Q20K2 (Q20) peptides to characterize the structural ensembles of these peptides. Classical MD simulations were performed on low energy structures taken from these metadynamics simulations to calculate equilibrium statistics on structures occupying energy minima.

UVRR experiments were also done on monomeric and fibril samples of Q15 and Q20 to calculate backbone Ψ dihedral angle distributions. Circular dichroism (CD) spectroscopy experiments were carried out to probe the secondary structure of these samples.

The Q15 peptide was solvated in 11 665 water molecules in a 72 × 72 × 72 Å³ box, and the Q20 peptide was solvated in 18 658 water molecules in a 84 × 84 × 84 Å³ box. These systems were prepared in VMD, and MD was simulated with NAMD. The CHARMM36 force field was used to calculate potential energies and forces, as it had been used successfully in our previous studies. TIP3P was used as the water model, and periodic boundary conditions were employed with a particle mesh Ewald grid spacing of 1.0 Å. The Verlet velocity integration scheme was used to move the atoms, and all simulations presented here were simulated under the NPT ensemble of constant temperature 300 K and pressure 1.013 25 bar. A nonbonded interaction cutoff of 12.0 Å was used with a pair-list distance of 14.0 Å and nonbonded switching distance of 10.0 Å.

We again used metadynamics to characterize the conformational free energy landscape for Q15 and Q20. Three RMSD coordinates were used as the collective variables that made up the conformational phase space for the energy landscape, as explained in Chapter 2. As previously, α-helix, β-hairpin, and PPII reference structures were used for these RMSD coordinates, and the RMSD measurement was conducted in
reference to the backbone atoms of the Q15 and Q20 peptides. The maximum allowable RMSD value for the coordinates was 15.0 Å and 20.0 Å for the Q15 and Q20 simulations respectively, and these limits were maintained with harmonic restraints of the form \( \frac{1}{2}k(x - x_0)^2 \) with a force constant of 1.0 kcal mol\(^{-1}\) Å\(^{-1}\). Gaussian energy hills of height 1.0 kcal mol\(^{-1}\) and width 3.0 Å were added to the potential energy of the simulation every 500 steps to explore the conformational energy landscape.

Each monomeric structure was minimized for 10,000 steps using the conjugate gradient minimization scheme\(^{43}\) and subsequently equilibrated at the correct temperature and pressure for 500 ps. Metadynamics were then simulated on the two peptide systems for a total of 400 ns each. This was considered to be sufficient sampling using metadynamics guidelines developed by Laio et al.,\(^{70}\) see Equation 1.11.

Following these simulations, analysis of the metadynamics data was performed using VMD\(^{83}\) along with custom Python scripts to locate low energy regions on the energy landscape and correlate those regions to sections of the MD trajectory. The structures occupying the local and global minima of the landscapes were then extracted using VMD.\(^{83}\) These structural states, two for Q15 and two for Q20, were then simulated using classical MD for a further 10 ns to gather equilibrium statistics for comparison to experiment. Potential energies and backbone Ψ dihedral angles were then obtained from the classical MD simulations using Tcl scripting in VMD.\(^{83}\) The R statistical package was used to visualize the landscapes.\(^{94}\)

### 5.3 Determination of Length Dependent Structural Properties for Q10, Q15, and Q20

Our collaborators find that Q15 can exist in two distinct structural states similar in structure to the PPII-rich (DQ10) and collapsed β-strand (NDQ10) states observed in Chapter 3. These are referred to as NDQ15 and DQ15; Ψ angle distributions for each of these states calculated from the UVRR spectra are shown in Figure 5.1a,b.

The Ψ distribution for NDQ15 shows a characteristic strong peak at 140° that
Figure 5.1: UVRR-obtained Ψ angle distributions of monomeric (a) NDQ15, (b) DQ15, and (c) DQ20. Computational Ψ angle distributions from metadynamics MD simulations for (d) Q15 β-strand, (e) Q15 PPII state, (f) Q20 PPII state, and (g) Q20 β-strand state. MD and UVRR distributions for monomeric Q10 can be found in Chapter 3. Source: figure prepared with Python package Matplotlib and Adobe Illustrator.
indicates the β-strand structure we previously observed in NDQ10. DQ10 yields a distribution (Figure 5.1c) that features a doublet peak with a maximum at 150°, corresponding to PPII structure, and shoulder peak located at 175°, indicative of the 2.5₁ type structure that we previously found was caused by the ionized aspartate and lysine terminal residues. Q20 was not able to be prepared in the collapsed β-strand state due to its marginal solubility, and UVRR spectra were only obtainable for the PPII state (referred to here as DQ20). The Ψ angle distribution calculated for this state is similar to the distribution for DQ15, with strong peaks that imply PPII and 2.5₁ structures. Interestingly, both DQ15 and DQ20 possess minor peaks around 0° that likely correspond to β-turn structural elements.

The results of the Q15 and Q20 conformational free energy landscapes from the metadynamics simulations are depicted in Figures 5.2 and 5.3.

Low-energy structures are identified by blue and purple colors on the energy landscape, as described in the legend next to each landscape. As was previously observed for Q10, the landscapes for Q15 and Q20 are frustrated, containing many shallow minima rather than small clearly defined energy wells. In both landscapes we observe two deep energy basins that are labeled Energy Well 1 and 2 on the landscapes. Representative structures populating these minima are displayed in Figure 5.2 with arrows indicating the wells they are located in. These structures were simulated with classical MD, and Ψ angle distributions were calculated. Notably, we find that the structures that populate the wells on the Q15 and Q20 landscape are relatively similar.

The Ψ angle distributions calculated from the MD simulations are shown in Figure 5.1d-g. Structures populating Energy Well 1 in both Q15 and Q20 have strong peaks at 140°, which is very close to the peak observed in the Ψ distribution for NDQ15, consistent with β-strand type structure. Because of this, we designate Energy Well 1 as the state corresponding to the collapsed β-strand conformation found
in the NDQ10 and NDQ15 peptides. For structures simulated from Energy Well 2, we find that the Ψ angle distribution has a peak at \(\sim 150^\circ\), which is consistent with a PPII-like conformation. There is also a minority population in the Ψ distribution centered at \(\sim 0^\circ\), characteristic of β-turn-like structure. The β-turn elements allow the PPII-rich structure to be collapsed for Q15 and Q20. The Ψ angle distributions for the Q15 and Q20 structures in Energy Well 2 in Figure 5.1 match the distributions observed for DQ15 and DQ20. Thus, we conclude that Energy Well 2 contains a predominately collapsed PPII-like structure interspersed with β-turn elements.
Figure 5.3: Conformational energy landscape for monomeric Q20 from metadynamics MD simulations. The lowest energy states are represented by the dark violet isosurfaces with example structures from the salient structural states shown. Source: figure prepared with R\textsuperscript{94} and Adobe Illustrator.

The Ψ angle distributions calculated from the UVRR spectra of DQ15 and DQ20 have peaks at ∼175° which are thought to be caused by 2.5₁-helix conformations. These peaks are not observed in the PPII-like structures from the metadynamics simulations. As discussed in Chapter 3, this discrepancy likely stems from an inadequacy in the CHARMM36 force field. Modern force fields are parameterized using X-ray and NMR structures of proteins, and 2.5₁-helix structures are uncommon structures that occur only in peptides with adjacent charged residues. Hence, it is unlikely that these structures would be favored by CHARMM36.

One obvious difference between the β-strand-like and PPII-like energy wells in
the metadynamics simulations is the breadth of the energy well. The β-strand-like energy well in the metadynamics simulations is conformationally broad, expanding a wide range of RMSD values. As a result, the metadynamics β-strand-like structure can vary with a low free energy cost. This suggests that the β-strand-like structure is flexible and may have significant conformational fluctuations. In contrast, the PPII-like energy well is narrow, indicating a more well-defined structure with less structural fluctuations.

Overall, our metadynamics simulations agree with our UVRR data showing that the structure of NDQ15 is predominately a β-strand-like conformation while the structure of DQ15 and DQ20 is predominantly a PPII-like conformation. This result is similar to that previously reported for DQ10 and NDQ10. We show that the experimental Ψ angle distributions for DQ15 and DQ20 quantitatively agree with that from the Q15 and Q20 PPII-like metadynamics structures. Similarly, the experimental Ψ angle distribution for NDQ15 quantitatively matches that from of the Q15 β-strand-like metadynamics structure.

The final aspect of this study was a CD spectroscopy analysis of the observed NDQ15, DQ15, and DQ20 structures. The CD spectra for these peptides is shown in Figure 5.4 and clearly indicate that DQ15 and DQ20 possess PPII secondary structure and NDQ15 has β-strand-like structure. However, compared to CD spectra collected for Q10, both the DQ15 and DQ20 spectra show a more negative and slightly redshifted peak at ~220 nm and slightly more positive peak at ~190 nm compared to DQ10. To probe these differences, our collaborators subtracted the DQ10 spectra from the spectra for DQ15 and DQ20. Interestingly, the difference spectrum contains a negative peak at ~220 nm and a stronger positive peak around 195 nm, which is characteristic of β-strand conformations and similar to the CD spectra for NDQ15. More so, the amount of β-strand character is greater in the spectra for DQ20 than the spectra for DQ15.
Figure 5.4: Solution-state CD spectra for DQ15, NDQ15, and DQ20 monomers. NDQ20 spectra was not obtainable because the peptide was not soluble. Source: figure prepared with Microsoft Excel.

This result indicates that residues in the PPII conformations (DQ15 and DQ20) have decreased PPII character and increased β-strand character, and we observe this trend increasing with polyQ repeat length. This conclusion was also drawn by Chellgren et al.,98 who found increasing β-strand character with increasing polyQ tract length.

The increased β-strand character in DQ15 and DQ20 peptides may suggest that longer polyQ peptides have a stronger preference for the β-strand-like structure compared to PPII structure. To investigate this, we examined the relative energies of the β-strand and PPII energy wells and associated energy barriers from the conformational free energy landscapes for Q15 and Q20. We also included data from the Q10 monomeric energy landscape calculated in Chapter 3.

Table 5.1 shows the relative Gibbs free energy minima of the β-strand-like and PPII-like energy wells from the metadynamics energy landscapes of Q10, Q15, and Q20. For each peptide, the PPII-like conformation has a higher energy compared to the β-strand conformation. We find that the PPII-like conformation increases in
energy with respect to the \( \beta \)-strand-like conformation as the polyQ repeat length increases. This result indicates that the PPII-like conformation is less energetically favorable compared to the \( \beta \)-strand conformation for longer polyQ peptides. This is in agreement with work by Darnell et al.,\textsuperscript{109} who used CD spectroscopy to probe the structure of polyQ peptides of increasing lengths. They observed PPII-like conformation decreasing and \( \beta \)-sheet content increasing with longer glutamine repeats. From this result they conclude that polyQ peptides are in a “tug of war” between PPII and \( \beta \)-strand structures, and as the repeat length increases this balance turns in favor of \( \beta \)-strand structures.

\[
\begin{array}{|c|c|c|}
\hline
\text{Peptide} & \beta\text{-strand} & \text{PPII-like} \\
\hline
Q10 & 0.0 & 0.0 \\
Q15 & 0.0 & 2.0 \\
Q20 & 0.0 & 2.0 \\
\hline
\end{array}
\]

**Table 5.1:** Relative free energy of lowest energy conformations for Q10, Q15, and Q20. Here, 0.0 represents the lowest \( \Delta G \) value.

Table 5.2 contains the various energy barriers between the PPII and \( \beta \)-strand states observed in their respective conformational energy landscapes. These barriers represent the lowest energy pathway between the energy minima. We find that the PPII to \( \beta \)-strand activation barriers are lower in comparison to the \( \beta \)-strand to PPII barriers for Q10, Q15, and Q20. Also, longer polyQ peptides have a lower activation barrier for the PPII to \( \beta \)-strand conversion. This suggests that longer polyQ peptides are increasingly capable of converting from PPII structures to \( \beta \)-strands, a conclusion that agrees with the proposition that longer polyQ peptides adopt \( \beta \)-strand prefer-

\[
\begin{array}{|c|c|c|}
\hline
\text{Peptide} & \beta\text{-strand to PPII} & \text{PPII to } \beta\text{-strand} \\
\hline
\text{\( \Delta G \) barrier (kcal mol\(^{-1}\))} & \text{\( \Delta G \) barrier (kcal mol\(^{-1}\))} \\
Q10 & 5.8 & 5.3 \\
Q15 & 5.2 & 3.1 \\
Q20 & 4.6 & 2.3 \\
\hline
\end{array}
\]

**Table 5.2:** Transition energy barriers between conformations of interest for Q10, Q15, and Q20. Here, 0.0 represents the lowest \( \Delta G \) value.
entially in comparison to PPII structures.

5.4 Conclusions and Direction

This work has shown that monomeric Q15 and Q20, like Q10, can form PPII-rich structures and \( \beta \)-strand-like conformations. We find that the PPII structures formed by longer polyQ peptides are more collapsed than the PPII conformations shown by Q10. This indicates that the monomeric polyQ ensemble that determines the first step of the aggregation process is indeed repeat length dependent. The real impact of this work was the observation from both CD spectroscopy and metadynamics that increasing polyQ repeat lengths preferentially adopt \( \beta \)-strand structures in comparison to PPII structures. This is supported by the energy landscapes, which show lower energy barriers for the PPII to \( \beta \)-strand transition as the number of glutamine repeats increases. An important question in the field is why longer polyQ peptides form cytotoxic aggregates when even short polyQ peptides (i.e. Q10) can aggregate into fibrils \textit{in vitro}. The work here suggests a framework for understanding why this is the case, as \( \beta \)-strand-rich peptides more rapidly form fibrils, and \( \beta \)-strand-rich peptides are favored in longer glutamine sequences.
Chapter 6. CONCLUSION

6.1 Summary of Dissertation Work

In this dissertation, we set out to deepen our understanding of the aggregation mechanism for polyQ peptides. The molecular mechanism of aggregation for polyQ-rich peptides is key to understanding and potentially treating polyQ diseases. Particularly, development of drugs that can bind to polyQ peptides and inhibit the aggregation process is dependent on knowledge of the different polyQ structures that exist along the aggregation mechanism. This dissertation sought to provide such knowledge by investigating the structure and thermodynamics of various polyQ states along the aggregation mechanism (example shown in Figure 1.3).

In the first study (Chapter 2), we used MD simulations along with UVRR spectroscopy experiments to investigate the structural properties of D$_2$Q$_{10}$K$_2$ (Q10) peptide fibrils. We were able to determine, by comparing Ψ dihedral angle distributions for three computational models of Q10 fibrils with distributions calculated from our collaborator’s UVRR spectra, that Q10 fibrils primarily adopt an extended, antiparallel β-sheet fibril structure which corresponded to the computational Model a. The fibril in Model a was observed to be stabilized by not only the backbone H-bonding typical for β-sheets, but also organized side chain H-bonding enabled by a side chain “zipper” conformation where the χ$_3$ side chain dihedral angle enables close packing of side chains. These results stand in contrast to studies of similar peptides which suggest they adopt β-hairpin conformations in fibrils. This conclusion establishes that extended antiparallel β-sheet structure is present in the aggregation mechanism for short polyQ peptides. The successful pairing of theory and experiment used in this study inspired us to use the same approach to study the monomeric structure of Q10.

Monomeric structure of polyQ peptides is an area of greater contention than that of polyQ fibrils. Using metadynamics MD simulations of aqueous Q10 peptides,
we were able to characterize the entire conformational free energy landscape of this peptide. We observed two deep conformational energy minima that corresponded to PPII-rich and collapsed $\beta$-strand structural states for Q10. Our collaborators found that they can poise Q10 to adopt two separate structural states, referred to as DQ10 and NDQ10. By analyzing and comparing distributions of backbone $\Psi$ dihedral angles calculated from both MD and UVRR, we determined that the experiment DQ10 peptide corresponded to the PPII-rich structural state from the MD and the NDQ10 peptide had the properties of the collapsed $\beta$-strand state. The conformational energy landscape we calculated with metadynamics indicated that these states were separated by a 3 to 6 $k_B T$ energy barrier. Experimentally, these peptide structural states were unable to interconvert at 300 K, confirming the presence of this energy barrier.

H-bonding analysis indicated that the collapsed $\beta$-strand conformational state was stabilized by an increased number of intrapeptide H-bonds, particular interactions between the side chain and backbone. Conversely, the PPII-rich state featured increased peptide$\leftrightarrow$water H-bonds and less intrapeptide interactions. The results from this study showed for the first time that monomeric polyQ peptides can populate distinct structural states; most literature in the past had concluded that polyQ monomers possess disordered structural ensembles.\textsuperscript{19,75,76,97} With these results, the initial steps of the polyQ aggregation mechanism are better understood. It is apparent that a $\beta$-hairpin aggregation scheme (seen on the right pathway in Figure 1.3) is not the correct mechanism for short polyQ peptide aggregation. Instead, the presence of a low energy $\beta$-strand monomeric structure gives credence to the type of aggregation mechanism shown on the left side of Figure 1.3. We set out next to better quantify the effect that H-bonding has on Q10 structure.

Although H-bonding of the glutamine side chain is reported to be a driving force of polyQ structure,\textsuperscript{32,76,77,115} no quantitative analyses of polyQ H-bonding strengths had been published. Here (Chapter 4), we strove to do this by again pairing the-
oretical and experimental methods. Our collaborators utilized the amide carbonyl band’s linear frequency dependence on H-bonding environment to develop a method to calculate a change in enthalpy of interaction (\(\Delta H_{\text{int}}\)) that occurs when a secondary amide interacts with a H-bonding environment. Using MD simulations, we are able to calculate an analogous value by computing the difference in energy between an interaction and a lack of interaction. We calculated experimental and theoretical \(\Delta H_{\text{int}}\) values for various types of side chain H-bonds in Q10 monomers and fibrils. \(\Delta H_{\text{int}}\) values calculated from MD simulations agreed nicely with those calculated from the UVRR spectra. This study showed that amideamide H-bonds formed by glutamine side chains to other side chains or backbone are stronger than H-bonds from polyQ side chains to water. An analysis of Q10 fibril side chain H-bonding demonstrated that the side chain H-bonds present in antiparallel \(\beta\)-sheet Q10 fibrils are enthalpically stronger than backbone H-bonds. These findings suggest that side chain H-bonding is an enthalpic driving force in the formation of collapsed polyQ peptide structures and an important stabilizing force in the formation of polyQ fibrils. Thus, polyQ structures that maximize peptide\(\leftrightarrow\)peptide H-bonding will be favored in the aggregation mechanism, and this may provide insight in determining unknown oligomeric polyQ structures. The last project we undertook was the application of some of the analyses from the studies of Q10 to peptides of longer sequence (Q15, Q20) to determine any repeat length dependent structural properties of polyQ peptides.

In Chapter 5 we used UVRR and metadynamics MD simulations to investigate the structural ensemble of \(D_2Q_{15}K_2\) (Q15) and \(D_2Q_{20}K_2\) (Q20) monomers. Metadynamics was used to calculate conformational free energy landscapes for these two peptides, and structures from the energy minima were simulated further with classical MD. UVRR and CD experiments were used to quantify backbone \(\Psi\) angle distributions and secondary structure content respectively. We observed that the conformational energy landscapes for Q15 and Q20 were relatively similar and both featured energy
wells corresponding to $\beta$-strand and PPII-rich structures. However, for these longer peptides the PPII-rich structures were collapsed, compared to the more extended PPII structures observed for Q10. Experimentally, it was found that Q15 could form DQ15 or NDQ15 peptides which correspond to the PPII and $\beta$-strand structures respectively. Due to its low solubility in water, Q20 was not able to be prepared in the NDQ20 state, so data for only DQ20 was collected. The NDQ$_N$ type structures corresponded to $\beta$-strand structures observed in the conformational energy landscapes, and the DQ$_N$ type structures corresponded to the PPII. The CD spectra echoed these results, but with an interesting observation: as the repeat length increased, more $\beta$-strand structure was observed in the PPII samples. $\beta$-strand monomers are known to form fibrils much more quickly than PPII monomers. Activation barriers and energy differences observed in the theoretical conformational landscapes strengthened these observations. As the glutamine repeat length increased, the PPII state relatively free energy rose compared to the $\beta$-strand. Also, as repeats increased the activation barrier for transition from PPII to $\beta$-strand decreased. These theoretical and experimental observations imply that increasing the length of polyQ repeats favors the formation of $\beta$-strand monomeric structures, which is a major finding in terms of the polyQ aggregation mechanism. This is an important finding, because it presents a framework for understanding why long polyQ repeats tend to aggregate more quickly in the cells of patients of polyQ diseases.

### 6.2 Future Directions and Unanswered Questions

Although we have uncovered interesting results in the four publications presented here, there is much more that needs to be understood in the field to build a comprehensive understanding of the aggregation mechanism of polyQ peptides, and thus the molecular pathology of polyQ diseases. Several questions that remain unresolved are posited here.

1. What are the morphologies of polyQ fibrils formed by peptides of longer repeat
lengths? Q10 has been shown to form fibrils that have a simple, relatively homogeneous antiparallel $\beta$-sheet structure. However, for longer polyQ peptides with a more flexible monomeric structural ensemble, this seems less likely. This might be approached by using MD simulations compared with UVRR, as we did for Q10 fibrils in Chapter 2, to determine the structural makeup of fibrils formed by longer polyQ fibrils. An appropriate choice of glutamine repeat length would be Q$_{36}$, as this is the critical length for the appearance of Huntington’s disease.\textsuperscript{10}

2. What role do aggregates and fibrils play in the pathology of polyQ diseases? Some studies have suggested that mature, insoluble fibrils actually play a beneficial role in the pathology of these diseases by sequestering smaller, soluble cytotoxic aggregates that disrupt function.\textsuperscript{21,95} However, this is not agreed upon throughout the field. Better understanding the cytotoxic components of these diseases would give biophysical researchers more specific species and mechanisms to study. Marsh et al.\textsuperscript{95} have established that cytosolic purely polyQ monomers are cytotoxic in \textit{Drosophila}. An interdisciplinary approach could be used, starting with experimental observation of a short sequence like Q10 in the cell. Potential sites of polyQ interactions in the cell could then be modeled with MD simulations to investigate the molecular mechanisms at play.

3. Although this work has established that increasing repeat lengths of polyQ sequences favor $\beta$-strand formation, this still leaves the question: what is unique about the repeat length that actually causes these diseases to manifest? This and other work has established that peptides well below physiological repeat lengths form amyloid-like fibrils, so what is special about the disease relevant repeat lengths? To learn this, a thorough investigation of repeat length dependent structural properties in polyQ monomers and fibrils is necessary, as well as an understanding of the relationship between repeat length and aggregation
mechanism. One such approach to the investigation of repeat length dependent structure is presented in Chapter 5 of this dissertation, however only three repeat lengths were used; this study would need to be expanded into the range of repeats that cause disease ($Q_{21-55}$). An approach to determining how the aggregation mechanism may be affected by increasing repeat length would require that various oligomeric states along the energy landscape (see Figure 1.3) be characterized. This could be accomplished by evaluating the free energy of formation for hypothesized oligomeric structures with MD simulations using enhanced sampling methods such as umbrella sampling$^{65}$ or metadynamics.$^{66}$
References


88. Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc.: Montreal, Canada., 2013.


