Insights into the Structure and Hydrogen Bonding Interactions of Polyglutamine Peptides Using UV Resonance Raman Spectroscopy

by

Ryan S. Jakubek

B.S. in Chemistry, University of Pittsburgh at Greensburg, 2013

Submitted to the Graduate Faculty of

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

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Ryan S. Jakubek

It was defended on

July 9, 2019

and approved by

Dr. Sunil Saxena, Professor, Department of Chemistry

Dr. Sean Garrett-Roe, Associate Professor, Department of Chemistry

Dr. Patrick van der Wel, Associate Professor, Zernike Institute for Advanced Materials

Dissertation Director: Dr. Sanford A. Asher, Distinguished Professor of Chemistry, Department of Chemistry

Insights into the Structure and Hydrogen Bonding Interactions of Polyglutamine Peptides Using UV Resonance Raman Spectroscopy

Ryan S. Jakubek, PhD

University of Pittsburgh, 2019

Proteins containing expanded polyglutamine (polyQ) tracts are prone to aggregation and fibrillization. This fibrillization is associated with many different neurodegenerative diseases including Huntington's disease. The pathology of these diseases is poorly understood. This is in part due to the difficulty of studying the structures and fibrillization mechanism of polyglutamine tracts. In this work, we utilize UV resonance Raman (UVRR) spectroscopy, as well as other techniques, to investigate the structures and hydrogen bonding interactions of solution-state, fibril-state, and insoluble forms of polyQ peptides. For example, we find that solution-state polyglutamine peptides can be prepared in an aggregation-prone β-strand-like conformation or an aggregation-resistant PPII-like conformation. We find that longer polyQ peptides have an increased population of aggregation-prone β-strand-like conformation that may contribute to faster aggregation kinetics and decreased age of onset for disease symptoms. We also developed a method to quantitatively correlate the frequency of the amide I (AmI) Raman band to the hydrogen bonding strength between the glutamine side chain C=O oxygen and a hydrogen bond donor. Using this technique, as well as other techniques, we investigated the oligomerization and hydrogen bonding interactions in the various structures of polyQ peptides to examine their role in polyQ aggregation.

Table of contents

List of tables

List of figures

List of equations

Preface

I would like to thank my advisor, Professor Sanford A. Asher, for his mentorship throughout my Ph.D. I greatly appreciate the freedom I was given to pursue my own research interests, make my own mistakes, and learn from the experience. My time in Dr. Asher's lab has allowed me to grow as a scientist and a person, and I could not be more grateful for the opportunities he has given me.

Throughout my Ph.D. I have collaborated with many wonderful scientists. I would like to thank David Punihaole for training and mentoring me during my first few years of graduate school. Without his influence I would not be where I am or who I am today. I would also like to thank Riley Workman for performing molecular dynamics and metadynamics simulations, and for insightful discussions that were crucial in understanding our research. I would also like to thank Stephen White for assistance in performing and analyzing experimental data and Krishnan Damodaran for preforming and helping to interpret NMR experiments. I would also like to thank all of my colleagues in Dr. Asher's research group for their support and friendship. I am especially grateful for the support of Kyle Hufziger, Ryan Roppel, Ivan Pallares, and Natasha Smith.

Finally, I would like to thank my friends and family for their support. I would like to thank my wife, Abby Tuttle, for her constant support. I would like to thank my dad and stepmom, Scott and Antonette Jakubek, for giving me security yet the space I needed to grow. And I would like to thank my mom, Myra Jakubek, for teaching me to never give up and for giving me a life-long love of learning. My friends and family have sacrificed so much to help me pursue my dream.

1.0 Introduction

1.1 Expanded Polyglutamine Tracts, Aggregation, and Neurodegeneration

1.1.1 CAG Codon Repeat Diseases

The expansion of CAG codon repeats in DNA encodes elongated polyglutamine (PolyQ) tracts in proteins. These elongated polyQ tracts induce protein aggregation that is linked to numerous neurodegenerative diseases, including Huntington's disease (HD).[1] A list of CAG repeat neurodegenerative diseases and their symptoms is provided in Table 1.1.

Table 1.1: CAG Repeat Expansion PolyQ Diseases

Disease	Affected Protein	Q repeat Disease Critical Length	Clinical Features
HD	Huntingtin	Q ≥ 36	Chorea, dystonia, cognitive defects, psychiatric problems
SCA1	Ataxin1	Q ≥ 39	Ataxia, Slurred speech, spasticity, cognitive impairments
SCA2	Ataxin2	Q ≥ 32	Ataxia, polyneuropathy, decreased reflexes, infantile variant with retinopathy
SCA3	Ataxin3	Q ≥ 61	Ataxia, parkinsonism, spasticity
SCA6	$CACNA1_A$	Q ≥ 10	Ataxia dysarthria, nystagmus, tremors
SCA7	Ataxin7	Q ≥ 37	Ataxia, blindness, cardiac failure in infants
SCA12	PPP2R2B	Q ≥ 55	Ataxia, head and hand tremor, akinesia
SCA17	TBP	Q ≥ 47	Ataxia, cognitive decline, seizures, psychiatric problems
SBMA	Androgen Receptor	Q ≥ 38	Motor weakness, swallowing, gynecomastia, decreased fertility
DRPLA	Atrophin	Q ≥ 49	Ataxia, seizures, choreoathetosis, dementia

HD, Huntington's disease; SCA, Spinocerebellar Ataxia; SBMA, Spinobulbar Muscular Atrophy; DRPLA, Dentatorubral-pallidoluysian dystrophy;

For CAG repeat diseases, longer polyQ tracts are correlated with an increase in neurodegenerative disease symptom severity and patient mortality.[1–5] For each disease, symptoms only occur if the polyQ tract surpasses a critical length. For example, Huntington's disease only occurs when the polyQ tract length in the huntingtin protein has ≥36 glutamine (Gln) residues.[1] In addition, longer polyQ tracts are correlated to a younger age of onset for disease symptoms.[5]

The presence of insoluble, polyQ-rich protein aggregates is observed in the neurons of patients suffering from CAG repeat diseases.[6,7] These aggregates generally appear as round or rod-shaped and are located in cell nuclei. The number of intranuclear inclusions is correlated with neuronal cell death, the severity of disease symptoms, and the length of the protein's polyQ tract.[7] Thus, polyQ-rich protein aggregates are thought to play an important role in CAG repeat disease pathology.

1.1.2 Toxicity of PolyQ in CAG Repeat Diseases

Much evidence suggests that CAG repeat diseases result from the protein gaining a toxic function as a result of the expanded polyQ tract. For example, Ordway et al. showed that expanded polyQ tracts will cause neurodegeneration and the formation of intraneuclear inclusions when included in a protein that is not naturally associated with CAG repeat diseases.[8] This suggests that toxicity originates from a gain-of-function from the expanded polyQ tract rather than a loss-of-function for a specific protein. In addition, it has been observed that polyQ peptides alone, outside of the context of a specific protein, can cause neurodegeneration.[9,10]

Though CAG repeat diseases are thought to result from protein gain-of-function, there is evidence to suggest that protein loss-of-function could also contribute to disease symptoms.

Studies have shown that wild-type huntingtin protects neuronal cells from various apoptotic stimuli.[11,12] This suggests that neurodegenerative disease symptoms for HD may partially result from loss-of-function of the huntingtin protein. In addition, Dragatsis et al. have shown that when huntingtin expression is disrupted, mice show neurological alterations similar to those in Huntington's disease.[13]

The toxic species of CAG repeat diseases is currently debated. The correlation between the presence of intranuclear inclusions and neuronal death suggests that the aggregate species may play a role in toxicity.[7] In addition, Yang et al. showed that synthetic, aggregated polyQ peptides are toxic to mammalian cells when introduced to the nuclei.[14] However, evidence also suggests that aggregation of polyQ-rich proteins may protect against neurodegeneration. For example, Arrasate et al. found that neuronal death is inversely correlated to the formation of Huntingtin aggregates.[15]

There is also evidence that monomeric proteins containing expanded polyQ tracts may be neurotoxic. For example, Nagai et al. found that a monomeric polyQ-rich protein species causes cytotoxicity when introduced to cultured cells.[16] Also, Trottier et al.[17] showed that monoclonal polyQ antibodies will recognize polyQ expansions in HD, Spinocerebellar Ataxia 1 (SCA1), and Spinocerebellar Ataxia 3 (SCA3), and that the binding affinity increases with polyQ length. From this they concluded that the antibody binds to a unique conformation of polyQ that is more prevalent for longer polyQ tracts. This study lead to the hypothesis that the toxic species might be a low concentration monomeric species that increases in concentration with polyQ tract length. However, the data by Trottier et al. could also result from the antibodies kinetically trapping polyQ conformations that are not otherwise populated. Also, the results of Trottier et al.

can be explained by the linear lattice effect, where increased binding of the antibody is due to an increase in epitopes for longer polyQ tracts.[18,19]

1.2 Protein Structure

The structure of a protein plays a large role in dictating its function. Thus, a detailed understanding of the pathology of CAG repeat diseases requires detailed knowledge of the structures of polyQ-rich proteins and peptides in their solution-state and aggregate forms. In this section we review the general motifs of protein structure.

Proteins are described as having four levels of structure: primary, secondary, tertiary, and quaternary structure (Figure 1.1).[20] The primary structure of a protein is the sequence of amino acids that comprise the protein.[20] Secondary structure refers to regular local conformations of the protein that are generally stabilized by short-range backbone-backbone hydrogen bonding.[20] Tertiary structure is a description of the completely folded protein that includes the packing of the secondary structural units by long-range interactions.[20] Finally, quaternary structure describes the structure of a protein that contains separate subunits that assemble.[20]

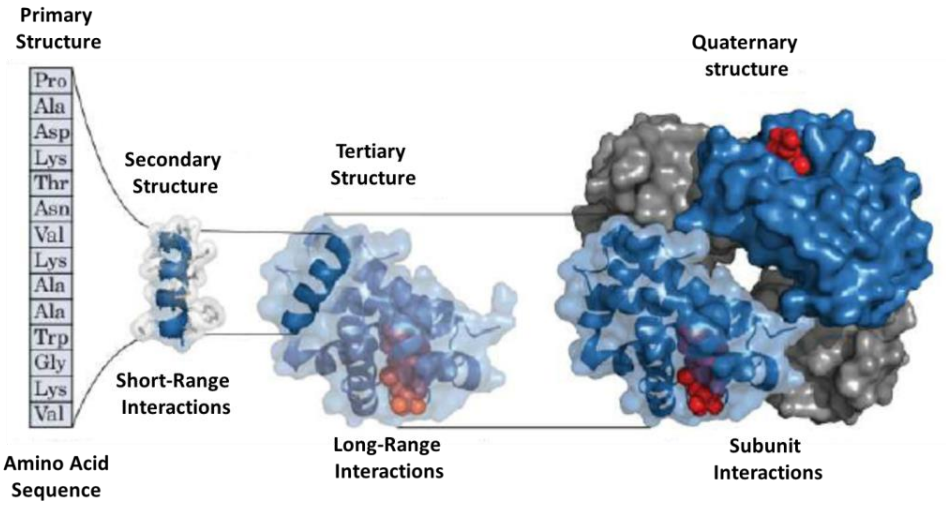

Figure 1.1: Depiction of the primary, secondary, tertiary, and quaternary levels of protein folding. Figure adapted with permission from Figure 3.23 from *Lehninger principles of biochemistry* 7ed by David L. Nelson and Michael Cox. Copyright 2017 by W.H. Freeman and Company. Used by permission of the publisher Macmillan Learning. Data from PDB ID 1HGA, R. Liddington et al., J. Mol. Biol. 228:551, 1992.

1.2.1 Peptide Backbone Torsion Angles

A protein or peptide's secondary structure is predominately defined by three torsion angles of the peptide backbone denoted Ψ (psi), Φ (phi), and ω (omega) (Figure 1.2).[21] The ω angle describes rotation around the C_i-N_{i+1} bond. Because the C_i-N_{i+1} bond has partial double bond character, the ω torsion angle is restricted to predominantly ~180°. Because the ω angle rotation is restricted, a peptide's secondary structure can be described by the Ψ and Φ torsion angles for each of its peptide bonds.[21]

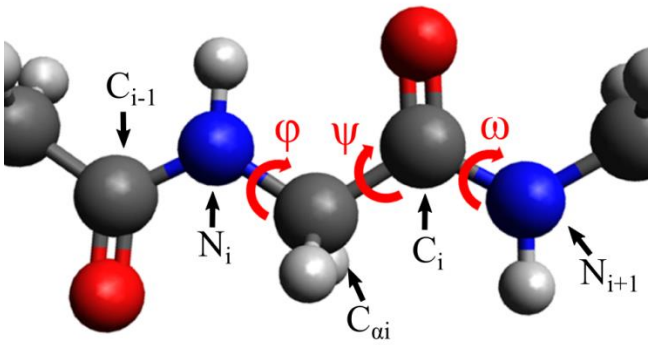

Figure 1.2: Peptide backbone showing the Ψ, Φ, and ω torsion angles.

The Ψ torsion angle describes rotation of the $C_{\alpha i}$-C_i bond and is defined as the dihedral angle between the $N_iC_{\alpha i}C_i$ and $C_{\alpha i}C_iN_{i+1}$ planes. The Φ torsion angle describes rotation of the N_i-$C_{\alpha i}$ bond and is defined as the dihedral angle between the $C_{i-1}N_iC_{\alpha i}$ and $N_iC_{\alpha i}C_i$ planes. Φ and Ψ torsion angles of 0° denote the cis conformation for the torsion angle. Clockwise rotation of the respective bond is denoted by positive angles while counterclockwise rotation is denoted by negative angles (Figure 1.3).

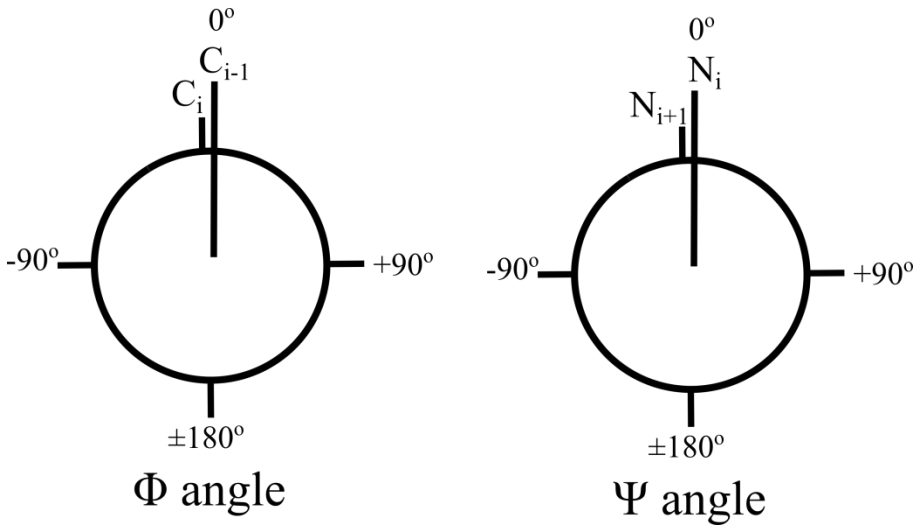

Figure 1.3: Neumann projections of the Φ and Ψ torsion angles showing the dependence of the angle on bond rotation.

A peptide's secondary structure can be visualized by plotting the backbone torsion angles on a Ramachandran plot where the Φ angle is the abscissa and the Ψ angle is the ordinate (Figure 1.4). As observed in Figure 1.4, certain regions of the Ramachandran plot are essentially forbidden due predominantly to steric restrictions.[21] Also, different protein secondary structures populate different regions of the Ramachandran plot. Thus, knowledge of a protein's Ramachandran Ψ and Φ angles allows for a detailed determination of its secondary structure.

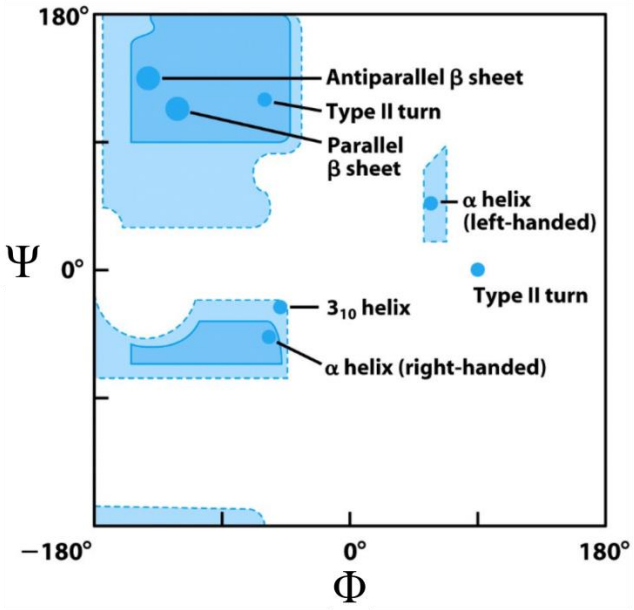

Figure 1.4: Ramachandran plot of common Ψ and Φ torsion angles. Dark blue shading indicates commonly populated regions while light blue shading indicates slightly less populated regions. White portions of the plot are rarely populated. The position of common secondary structures is indicated on the Ramachandran plot. Horton, Robert A; Moran, Laurence A.; Scrimgeour, Gray; Perry, Marc; Rawn, David, Principles Of Biochemistry, 4th, ©2006. Reprinted by permission of Pearson Education, Inc., New York.

1.2.2 Amino Acid Side Chain Structure

The amino acid side chain structures can also be defined by the torsion angles of rotatable bonds that comprise the side chain. The side chain torsion angles are denoted as χ_1, χ_2, χ_3, etc. where χ_1 describes rotation of the C_α-C_β bond, χ_2 describes rotation of the C_β-C_γ bond, and so on (Figure 1.5). The possible conformations of side chains are dependent on the peptide backbone secondary structure with different secondary structures sterically forbidding certain side chain conformations.[22,23]

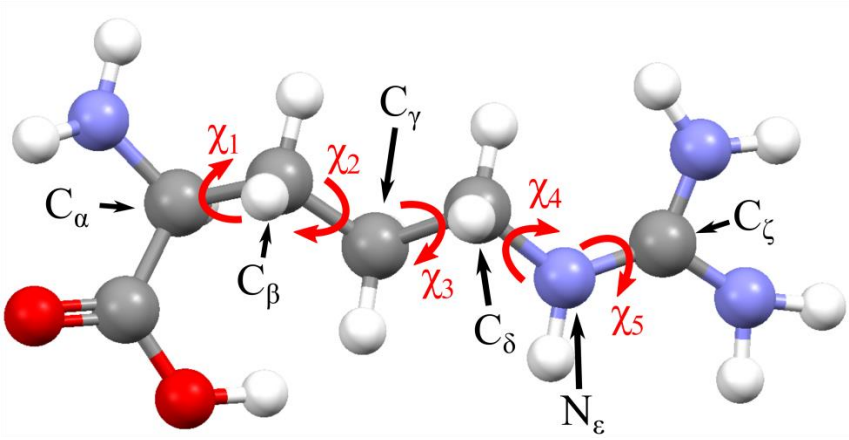

Figure 1.5: Arginine amino acid with side chain atoms and torsion angle labeled.

1.3 Secondary Structure Motifs

1.3.1 Helix Structures

There are a variety of helical peptide backbone conformations each with a different number of amino acids per helical turn (Figure 1.6a). The α-helix is a right handed helix that contains 3.6 amino acid residues per turn of the helix and has characteristic backbone (Φ,Ψ) angles of (-57°,-47°).[21,24,25] α-helix structures are stabilized by hydrogen bonding interactions between the C=O group of every i^{th} residue with the N-H group of every $i+4^{th}$ residue.[21]

In contrast, the 3_{10}-helix (Figure 1.6b) contains 3 residues per helical twist with characteristic (Φ,Ψ) angles of about (-49°,-26°) in peptides.[21,24,25] The 3_{10}-helix is stabilized by the i^{th} residue backbone C=O hydrogen bonding to the $i+3^{rd}$ N-H group.[21] The backbone hydrogen bonding in the 3_{10}-helix is weaker than that of the α-helix.[26] As a result, the 3_{10}-helix conformation is less stable and less common compared to α-helices.[26,27]

Another helical structure is the π-helix (Figure 1.6c),which consist of a right handed helix with (Φ,Ψ) angles of (-57°,-70°).[24,25] The π-helix contains 4.4 residues per helical turn and is stabilized by hydrogen bonding between the C=O group of the i^{th} residue and the N-H group of the $i+5^{th}$ residue of the helix.[26] As with the 3_{10}-helix, the π-helix is less common than the α-helix because the hydrogen bonding interactions stabilizing the π-helix are less favorable.[26]

The polyproline II (PPII) -helix (Figure 1.6d)contains (Φ,Ψ) angles of (-75°,150°).[28] The PPII-helix conformation does not allow for the formation of backbone-backbone hydrogen bonds and has a large exposed surface allowing for extensive water-peptide hydrogen bonding.[29]

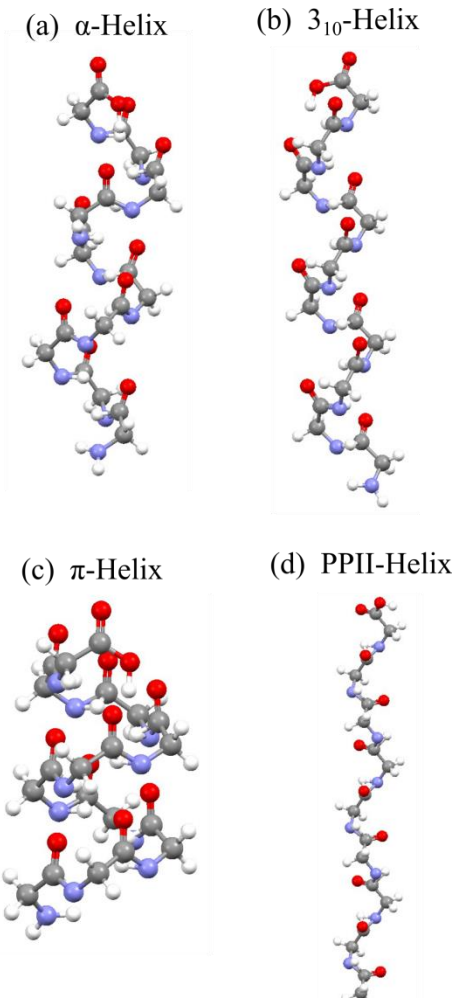

Figure 1.6: Backbone structure of an (a) α-helix, (b) 3_{10}-helix, (c) π-helix, and (d) PPII-helix.

1.3.2 β-Sheet and β-Strand Structures

β-strand structures are extended peptide chain conformations. Multiple adjacent β-strands can form inter-β-strand hydrogen bonds to form a β-sheet. Adjacent β-strands within a β-sheet can either be oriented in the same N- to C- terminal direction, parallel β-sheet (Figure 1.7a), or in the opposite N- to C- terminal direction, antiparallel β-sheet (Figure 1.7b). The inter-β-strand hydrogen bonds in parallel β-sheets are weaker compared to that in the antiparallel β-sheets. As a result the antiparallel β-sheet structure is more stable compared to the parallel β-sheet.[21] The Φ and Ψ angles in β-sheets occupy a broad region of the Ramachandran plot.[21] Antiparallel β-sheets have (Φ,Ψ) angles found at about (-139°,135°) while that of parallel β-sheets have (Φ,Ψ) angles found at about (-119°,113°).[24,25]

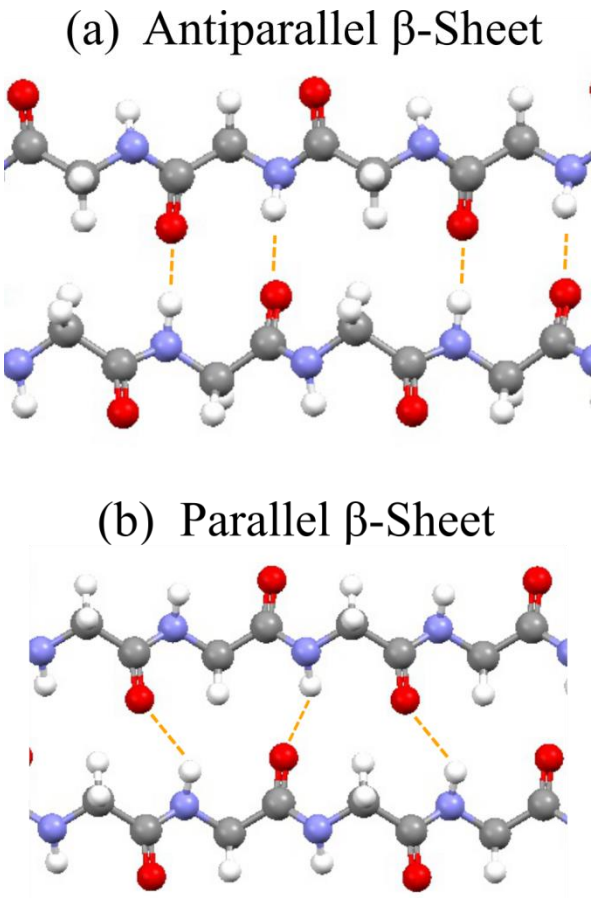

Figure 1.7: Structure of (a) antiparallel and (b) parallel β-sheets. The dotted lines show inter-β-strand hydrogen bonding.

1.3.3 Intrinsically Disordered Proteins

Intrinsically disordered proteins (IDP) or protein domains contain little secondary and tertiary structure. Instead they have a flexible, random-coil-like behavior that is thought to play an important role in their biological functions.[30]

1.4 Polyglutamine Peptide Structure

1.4.1 Solubilization of PolyQ Peptides

In general polyQ peptides have low solubility in water.[31] The solubility of polyQ peptides can be increased by adding polar amino acids to the C- and N-termini of polyQ tracts[32], however, even with the addition of polar residues, long polyQ tracts (Q>~20) are generally observed to be insoluble in water.[31] This severely hinders experimental investigations of solution-state polyQ peptide structures and their aggregation mechanism(s).

To circumvent this, Chen et al. developed a procedure to solubilize polyQ peptides.[31] This procedure involves dissolving and incubating polyQ in a 1:1 hexafluoroisopropanol (HFIP) trifluoroacetic acid (TFA) mixture. After 0.5-4 hours of incubation at room temperature, the HFIP/TFA solvent is evaporated, and the peptide is re-dissolved in water or buffer. This procedure is known as disaggregation because the procedure is thought to remove the presence

of small aggregates that seed aggregation without affecting the structure of the peptide.[33] However, the mechanism by which disaggregation solubilizes polyQ peptide is not completely understood.

1.4.2 Proposed Solution-State Structures

Experimental[34–36] and computational[37–39] studies suggest that polyQ peptides are predominantly disordered in solution. However, there is evidence that polyQ tracts can adopt transient secondary structures that may play a role in polyQ aggregation.[16,40] The structure of solution-state polyQ peptides was found to be similar, regardless of polyQ tract length.[37,38,41,42] This includes polyQ tract lengths below and above the pathogenic threshold.

The disordered structure of polyQ tracts was found to significantly differ from a true random coil polymer.[37] Experimental studies have shown that longer polyQ peptides have a more collapsed structure.[43,44] It is thought that this is caused by a decrease in the water solvent quality for longer polyQ peptides with water being a good solvent for Q<16, a theta solvent for Q=16, and a poor solvent for Q>16.[43–45] The decrease in solvent quality with increasing polyQ length is thought to drive the formation of inter- and intra-peptide hydrogen bonds promoting polyQ aggregation.

1.4.3 Proposed Fibril Structures

PolyQ peptide aggregates grown *in vitro* exhibit many amyloid-like features, including fibril-like morphologies in electron microscopy (EM) images,[33,46] β-sheet structural features,[47–50] thioflavin T (ThT) binding,[35] and cross-β-like X-ray diffraction patterns.[32,46,49] Early structural

models of polyQ fibrils were based on X-ray diffraction data. Because single crystalline polyQ fibrils could not be prepared, X-ray diffraction cannot provide atomic resolution structures. As a result, the structural models derived from X-ray powder patterns can be interpreted in multiple ways.

In 1994, from x-ray diffraction data, Max Perutz and coworkers proposed a polar zipper structure for fibrils grown from the peptide $D_2Q_{15}K_2$.[32] In this model polyQ forms a cross β-structure composed of β-sheets, where the β-strand direction is perpendicular to the long axis of the fibril. However, Perutz et al.[51] later proposed a water-filled nanotube structural model from the same x-ray diffraction data (Figure 1.8a). The water-filled nanotube structure consists of a cylindrical β-sheet structure, where successive turns in the cylinder are stabilized by side chain-side chain and backbone-backbone hydrogen bonds. Sikorski and Atkins[49] later reinterpreted the X-ray diffraction data of Perutz et al.[32] and proposed a cross-β fibril structure that has a different structure and hydrogen bonding interactions compared to the models proposed by Perutz et al. (Figure 1.8b).[32,51]

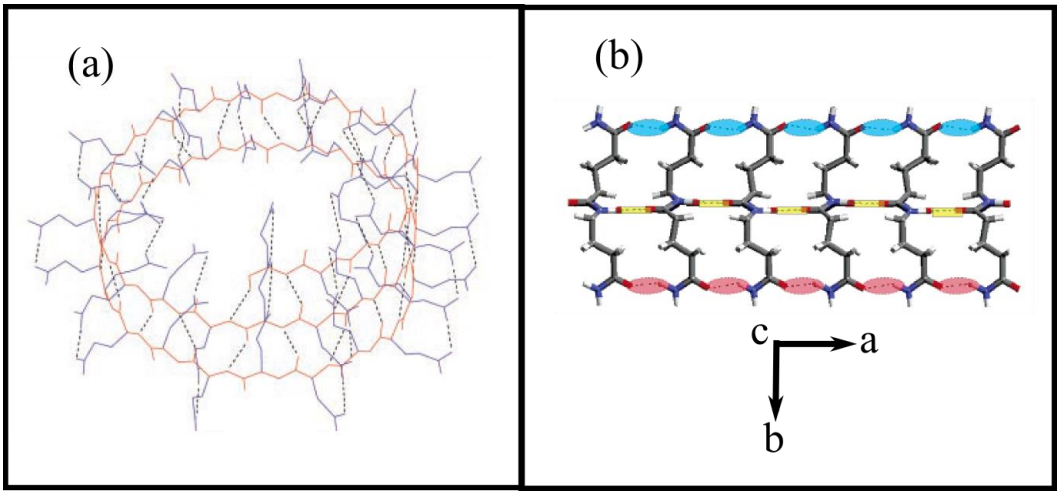

Figure 1.8: PolyQ fibril models. (a) water-filled nanotube model of polyQ fibrils developed by Perutz et al.[51] The red line indicates the peptide backbone, the black lines indicate Gln side chains, and the dotted lines indicate hydrogen bonds. (b) cross-β structural model developed by Sikorski and Atkins.[49] In (b), the "a" axis indicates the direction of the fibril axis, the "b" axis is orthogonal to the fibril axis and β-sheet plane, and the "c" axis is parallel to the β-strand direction. The blue and red highlighting indicate side chain-side chain hydrogen bonds and the yellow highlighting indicate the canonical backbone-backbone hydrogen bonding found in a β-sheet. Figure (a) adapted with permission from Perutz M. F., et al. *PNAS* 2002, 99 (8), 5591–5595. Copyright 2002 Proceedings of the National Academy of Sciences of the United States of America. Figure (b) adapted with permission from Sikorski P. et al. *Biomacromolecules.* 2005, 6 (1), 425–432. Copyright 2005 American Chemical Society.

In addition, various other experimental techniques, such as NMR,[50,52–54] IR,[55,56] and UV resonance Raman[46,57] (UVRR) spectroscopies, have been used to obtain detailed information on the polyQ fibril structure. These spectroscopic studies have investigated many important structural aspects of polyQ fibrils, such as, for example, the structure of the Gln side chains[46,52] and the structure of fibril monomer unit.[46,52,55]

1.4.4 PolyQ Aggregation

Wetzel and coworkers have intensively investigated the kinetics and mechanism of polyQ peptide aggregation.[33,35,48,58–60] They found that polyQ peptide aggregation consists of an initial lag phase before aggregation begins.[33] This is consistent with a nucleation-dependent aggregation process. Longer polyQ peptides were found to have shorter lag phases and faster aggregation kinetics.[33] However, when seeded with pre-formed polyQ aggregates, no lag phase is observed.[33] From these data they propose that polyQ aggregation occurs via a nucleation-dependent aggregation mechanism, where the nucleus is a thermodynamically unfavorable polyQ conformation, such as a β-hairpin structure (Figure 1.9).[58] Kar et al.[58] proposes that the nucleus size for polyQ aggregation is dependent on the peptide length, with Q>25 having monomeric nuclei and Q<25 having oligomeric nuclei. However, monomeric nuclei are uncommon in nucleation-dependent aggregation processes.[35]

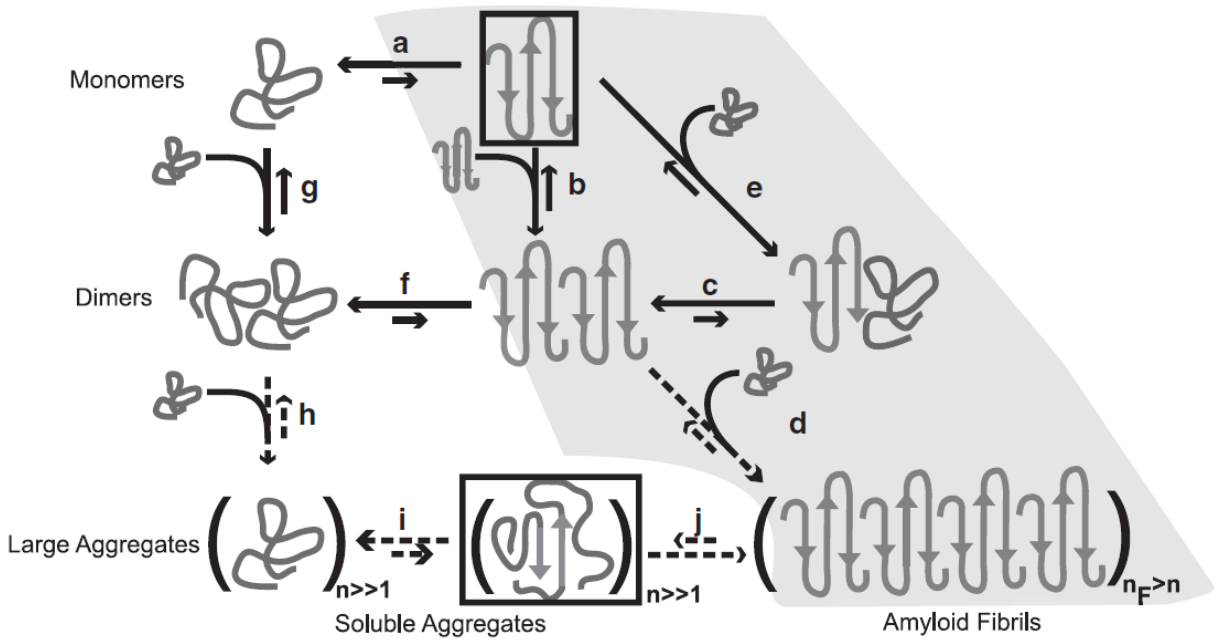

Figure 1.9: Possible aggregation mechanisms of polyQ peptides. The mechanisms shaded in gray are those proposed by the Wetzel group, where aggregation occurs through a monomeric nucleus and monomer is then recruited onto the nucleus or growing fibril. The mechanism in white is that proposed by the Pappu group, where polyQ peptides aggregate into unordered oligomers that then convert to fibrils. Reprinted with permission from Vitalis A., et al. *Biophys. J.* 2009, 97 (1), 303–311. Copyright 2015 American Chemical Society.

In contrast, Pappu and coworkers have used computational methods to propose an alternative mechanism for polyQ fibrillization.[37,38,45,61] They argue that polyQ peptides form disordered, high molecular weight oligomers that convert into fibrils (Figure 1.9).[61] Both computational[37,38] and experimental[43,44] results indicate that longer polyQ peptides increasingly favor peptide-peptide interactions over peptide-water interactions. Water was found to be a good solvent for Q<16, a theta solvent for Q=16, and a poor solvent for Q>16.[43–45] The preference for peptide-peptide interactions in long polyQ peptides is thought to drive polyQ aggregation and eventually fibril formation.[43,45]

1.5 The need for Novel Techniques to Investigate Polyglutamine Structure

As shown above, there is a lack of consensus in the current understanding of polyQ peptide solution-state structure, fibril structure, and aggregation mechanism. Numerous techniques, such as NMR, IR, and X-ray crystallography, have provided great insight into the structures of polyQ peptides. Each technique provides a unique way of investigating polyQ structure and elucidates different aspects of the polyQ solution and/or fibril structures. Further insights into the structures of polyQ peptides necessitate the development and use of new techniques that can investigate different structural aspects polyQ peptides.

UVRR spectroscopy is a relatively new biophysical tool that can elucidate the structure proteins and peptides.[62] Recently the Asher group has characterized a variety of UVRR spectral markers that are sensitive to peptide structure and hydration.[63] This includes the structure and

hydration of the peptide backbone[64,65] as well as a variety of amino acid side chains.[66–72] These spectroscopic markers have been used to probe peptide and protein structure[73–76] including recent work investigating the structure of polyQ peptides.[46,57] The work described in this dissertation continues to develop and use UVRR spectroscopy as a tool for the investigation of polyQ structure.

2.0 Raman Scattering

When light is incident on a molecule, a fraction of that light is scattered. Most of the scattered light is at the same frequency as the incident light. This elastic scattering is referred to as Rayleigh scattering. In contrast, Raman scattering is an inelastic light scattering phenomenon where the scattered light is frequency shifted with respect to the incident light. The frequency shift of the Raman scattered light is equal to the frequency of a molecular vibration. Therefore, the spectrum of Raman scattered light provides a vibrational spectrum of the analyte of interest. Here we give a brief discussion on Raman theory topics that are important to understand this work. For a complete description of Raman theory we direct you to the work of Derek A. Long.[77]

2.1 Classical Theory of Rayleigh and Raman Scattering

In Rayleigh and Raman light scattering, incident light induces an oscillation of a molecule's electrons, producing an induced oscillating dipole. This induced oscillating dipole is the source of the scattered light. The intensity of radiation produced from an oscillating electric dipole is given by:

$$I = k\,\omega_s^4\,p_0^2\,sin^2\theta \qquad \textbf{2.1}$$

where I is the intensity of the scattered light, k is a constant, ω_s is the frequency of the scattered light, p_0 is the amplitude of the dipole moment, and θ is the angle between the scattered light propagation vector and the dipole moment vector. The dipole moment amplitude (p_0) is the product of the molecule's polarizability (α) and the amplitude of the electric field (E_0) produced by the incident light:

$$p_0 = \alpha E_0 \qquad \textbf{2.2}$$

where E_0 is the electric field vector of coherent radiation of frequency ω_1, and α is the molecule's polarizability. The molecule's polarizability is predominantly constant, resulting in an induced electric dipole (p_0) that oscillates at the same frequency as the incident light (ω_s). This results in scattered light at the incident frequency (ω_1). However, the polarizability of the electrons in the molecule is dependent on the molecule's normal coordinates. As the molecule vibrates, its normal coordinates and polarizability changes. This produces a component of the dipole (p_0) that oscillates at a frequency different from that of the incident light. This component of the dipole produces Raman scattered light where the frequency shift of the light is equal to the frequency of the molecular vibration ($\omega_1 \pm \omega_k$).

The variation of polarizability with a molecule's vibration can be expressed by expanding the polarizability in a Taylor series with respect to the vibrational normal coordinate (Q_k):

$$\alpha = \alpha_0 + \sum_k \left(\frac{\partial \alpha}{\partial Q_k}\right)_0 Q_k \ldots \qquad \textbf{2.3}$$

where the "0" subscript denotes that the value is taken at the molecule's equilibrium configuration, and k is a given normal mode of the molecule. Equation 2.3 can be written as:

$$\alpha = \alpha_0 + \alpha'_k Q_k \qquad \textbf{2.4}$$

where

$$\alpha'_k = \sum_k \left(\frac{\partial \alpha}{\partial Q_k}\right)_0 \qquad \textbf{2.5}$$

The time dependence of the vibrational normal coordinate (Q_k) is given by equation 2.6 assuming simple harmonic motion:

$$Q_k = Q_{k0} \cos(\omega_k t + \delta_k) \qquad \textbf{2.6}$$

where Q_{k0} is the vibrational amplitude, ω_k is the vibrational frequency, and δ_k is a phase factor. Combining eqs. 2.5 and 2.6 we obtain the time dependence of the polarizability with respect to the k^{th} normal mode.

$$\alpha = \alpha_0 + \alpha'_k Q_{k0} \cos(\omega_k t + \delta_k) \qquad \textbf{2.7}$$

The incident radiation's electric field oscillates at frequency ω_1:

$$E = E_0 \cos(\omega_1 t) \qquad \textbf{2.8}$$

where E_0 is the electric field amplitude. Thus,

$$p = \alpha E = \alpha_0 E_0 \cos(\omega_1 t) + \alpha'_k Q_{k0} E_0 \cos(\omega_k t + \delta_k) \cos(\omega_1 t) \qquad \textbf{2.9}$$

The first term of 2.9 is a component of the dipole that oscillates at the incident light frequency (ω_1), giving rise to Rayleigh scattering. In the second term the dipole oscillates at a frequency dependent on the incident light frequency and kth normal mode frequency, giving rise to Raman scattering. It is also important to note that the magnitude of the second term in eq. 2.9, and thus the Raman scattering intensity, is dependent on the change in polarizability with respect to the normal coordinate at the equilibrium geometry ($\alpha'_k Q_{k0}$). This is referred to as the Raman polarizability. The second term in 2.9 can be reformulated using the following identity.

$$\cos(A)\cos(B) = 1/2(\cos(A + B) + \cos(A - B)) \qquad \textbf{2.10}$$

Thus

$$p = \alpha_0 E_0 \cos(\omega_1 t) + \alpha'_k Q_{k0} E_0 \cos(\omega_1 t - \omega_k t + \delta_k) + \alpha'_k Q_{k0} E_0 \cos(\omega_1 t + \omega_k t + \delta_k) \qquad \textbf{2.11}$$

Equation 2.11 shows that the Raman component of the induced dipole oscillates at the frequencies of ω_1-ω_k and ω_1+ω_k corresponding to Stokes and anti-Stokes Raman scattered light

respectively. Equation 2.11 shows that the change in the polarizability of the molecule with respect to a normal mode of vibration gives rise to Raman scattered light.

The classical description of Raman scattering is depicted in Figure 2.1. The frequency of the incident light (a) drives the oscillation of electrons in a molecule that radiate light. For the electric field interacting with the component of the polarizability that is constant (b), the resulting dipole oscillates at and emits radiation at the incident light frequency (c). In contrast, for an electric field interacting with a polarizability oscillating at the frequency of a molecular vibration (d), the resulting dipole oscillates as a beat frequency (e) with ω_1, $\omega_1-\omega_k$, and $\omega_1+\omega_k$ frequency components (f).

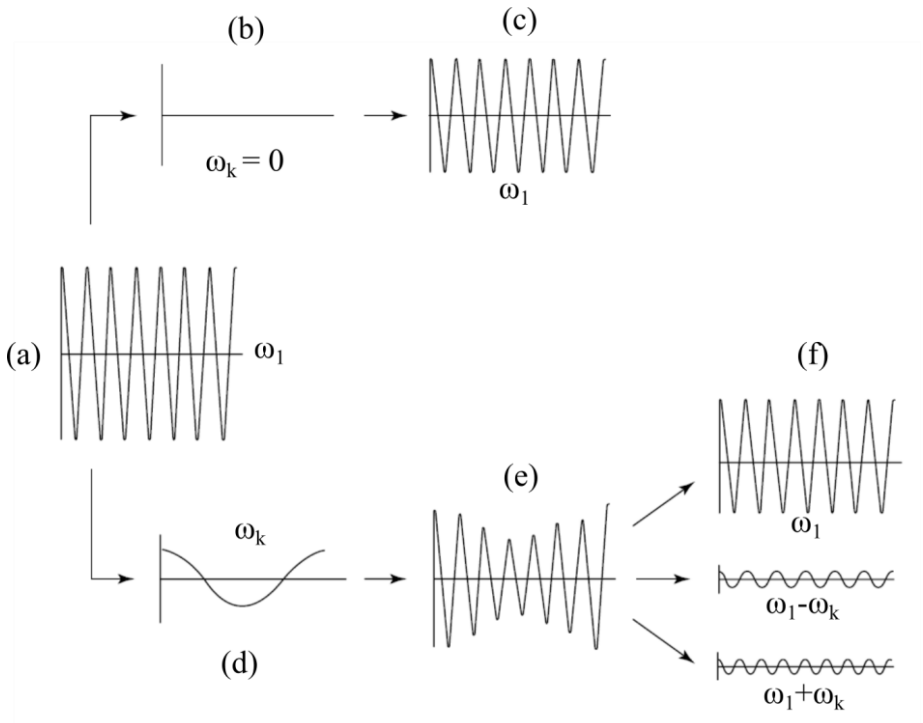

Figure 2.1: Effect of an oscillating molecular polarizability on the electric dipole induced by incident radiation. (see text for details) Adapted with permission from Long, D. A. The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules; John Wiley and Sons Ltd.: Chichester, 2002. Copyright 2002 John Wiley and Sons Ltd..

2.2 Quantum Mechanical Treatment of Rayleigh And Raman Scattering

The existence and frequency shifts involved in Raman scattering are correctly predicted from classical physics. However, to qualitatively understand scattering intensities we must obtain information on the molecule's polarizability, which invokes a quantum mechanical treatment of Raman scattering. The quantum mechanical treatment of Raman theory treats the scattering molecule quantum mechanically and the radiation with classical physics.

The processes of Rayleigh and Raman scattering can be visualized via the energy transfer model. Figure 2.2 depicts the energy transfers between the incident radiation and the molecule that occur in Rayleigh and Raman scattering. Rayleigh and Raman scattering can be thought of as a simultaneous absorption and emission processes. In non-resonance Raman and Rayleigh scattering, the incident light is not in resonance with an electronic transition in the molecule. Therefore, in the non-resonance case, the molecule undergoes virtual absorption to a virtual excited state that emits a photon as the molecule relaxes to a discrete quantum state of the system. Virtual absorption is absorption without the conservation of energy that excites to a virtual state of the system, which is not an eigenstate of the system.

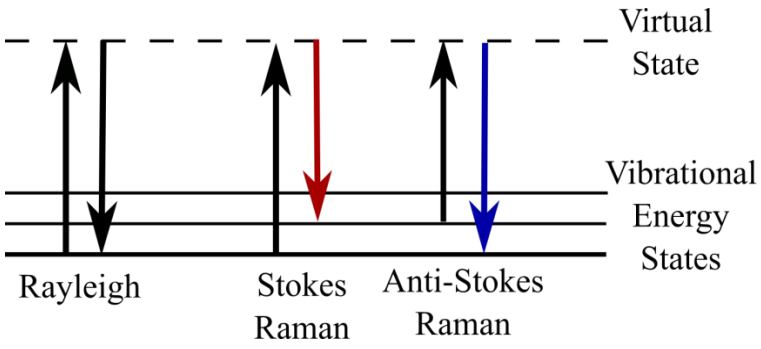

Figure 2.2: Energy level diagram depicting the energy transfer that occurs for Rayleigh, Stokes Raman, and Anti-Stokes Raman processes.

In Rayleigh scattering there is no transfer of energy between the molecule and the photon. A photon is virtually absorbed by the molecule, and the excited molecule emits a photon of the same frequency as it returns to the ground vibronic state. In contrast, in Raman scattering there is an energy exchange between an incident photon and the molecule. In Stokes Raman scattering, the incident photon is virtually absorbed, and the molecule relaxes to an excited vibrational state of the ground electronic state. As a result the emitted photon has a lower energy compared to the incident photon. The difference in energy between the incident photon and the Raman scattered photon corresponds to the energy difference between the initial (ground) and final (excited) vibrational energy level of the molecule. Therefore, the frequency shift of the Raman scattered light corresponds to the frequency of the molecular vibration (ω_k) being excited. In anti-Stokes Raman scattering, the initial state of the molecule is an excited vibrational state. The molecule then virtually absorbs a photon and emits a photon as the molecule relaxes to a vibrational energy level lower than its initial vibrational energy level. The resulting scattered

photon has a higher energy, and frequency, compared to the incident photon. Below is a brief mathematical description of the quantum theory of Raman scattering following the work of Long.[77]

In the quantum physics model, the classical dipole moment is replaced by the transition electric dipole (P_{fi}) that induces the transition of a molecule from state *i* to state *f*.

$$P_{fi} = \langle \psi'_f | P | \psi'_i \rangle \qquad \textbf{2.12}$$

Here, Ψ' are the molecule's time-dependent perturbed wave functions for state *i* and *f* and can be represented by the following series expansions:

$$\psi'_f = \psi_f^{(0)} + \psi_f^{(1)} + \cdots \psi_f^{(n)} \qquad \textbf{2.13}$$

and

$$\psi'_i = \psi_i^{(0)} + \psi_i^{(1)} + \cdots \psi_i^{(n)} \qquad \textbf{2.14}$$

where $\psi^{(0)}_i$ is the unperturbed state, and $\psi^{(n)}_i$ is the n^{th} order modification to the unperturbed state.

By substituting eqs. 2.13 and 2.14 into eq. 2.12, we obtain expressions for the n^{th} order modifications to the transition dipole moment.

$$P^{(0)}_{fi} = \left\langle \psi_f^{(0)} \middle| P \middle| \psi_i^{(0)} \right\rangle \qquad \textbf{2.15}$$

$$P^{(1)}{}_{fi} = \langle \psi_f{}^{(1)} | P | \psi_i{}^{(0)} \rangle + \langle \psi_f{}^{(0)} | P | \psi_i{}^{(1)} \rangle \quad \textbf{2.16}$$

$$P^{(2)}{}_{fi} = \langle \psi_f{}^{(0)} | P | \psi_i{}^{(2)} \rangle + \langle \psi_f{}^{(2)} | P | \psi_i{}^{(0)} \rangle + \langle \psi_f{}^{(1)} | P | \psi_i{}^{(1)} \rangle \quad \textbf{2.17}$$

The first-order perturbed dipole moment (eq. 2.16) describes the Raman scattering phenomenon. The procedure for evaluating eq. 2.16 is outlined by Long[77]. Briefly, a relationship is obtained between the perturbed time-dependent wave functions and unperturbed time-dependent wave functions assuming that the perturbation is entirely electric dipole in nature and is induced entirely by the incident electric field. The relationships are then substituted in eq. 2.16 for the perturbed time dependent wave functions. The Raman component of the resulting equation is as follows:

$$(P_\rho^{(1)})_{fi} = \frac{1}{2\hbar} \sum_{r \neq i,f} \left\{ \frac{\langle f | \hat{p}_\rho | r \rangle \langle r | \hat{p}_\sigma | i \rangle}{\omega_{ri} - \omega_1 - i\Gamma_r} + \frac{\langle f | \hat{p}_\sigma | r \rangle \langle r | \hat{p}_\rho | i \rangle}{\omega_{rf} + \omega_1 + i\Gamma_r} \right\} E_{\sigma 0} \exp(-i\omega_s t) + C.C. \quad \textbf{2.18}$$

where, $\omega_{ri} = \omega_r - \omega_i$, $\omega_{rf} = \omega_r - \omega_f$, $\omega_s = \omega_1 - \omega_{fi}$, C.C. is the complex conjugate, $E_{\sigma 0}$ is the σ component of the complex amplitude of the incident radiation electric field, and Γ_r is the damping factor of eigenstate (r) which is related to the full width and lifetime of state r. The equation is summed

over all states r of the molecule. We have simplified the notation by substituting ψ_i for i and so on.

Because $P = \alpha E$, we can define the Raman transition polarizability as:

$$(\alpha_{\rho\sigma})_{fi} = \frac{1}{\hbar}\sum_{r\neq i,f}\left\{\frac{\langle f|\hat{p}_\rho|r\rangle\langle r|\hat{p}_\sigma|i\rangle}{\omega_{ri} - \omega_1 - i\Gamma_r} + \frac{\langle f|\hat{p}_\sigma|r\rangle\langle r|\hat{p}_\rho|i\rangle}{\omega_{rf} + \omega_1 + i\Gamma_r}\right\} \qquad \textbf{2.19}$$

For an induced oscillating dipole driven by incident light with a constant frequency and electromagnetic field amplitude, the intensity of the scattered light monitored at a given scattering angle from the dipole axis is proportional to the Raman transition polarizability squared (eq. 2.1). Thus eq. 2.19 provides insight into the intensity of Raman scattered light.

2.3 Resonance Raman Scattering

In resonance Raman scattering the incident light frequency is in resonance with an electronic transition. As a result, the initial (i) and final (f) states of the Raman process are in the ground electronic state and the transition state (r) is an excited vibronic eigenstate of the system (Figure 2.3). By invoking the Born-Oppenheimer approximation and neglecting Herzberg-Teller vibronic coupling, we obtain the following equation for the transition Raman polarizability:

$$(\alpha_{\rho\sigma})_{e^f v^f : e^g v^i} = \frac{1}{\hbar} \sum_{e^r v^r \neq e^f v^f : e^g v^i} \left\{ \frac{\langle v^f | (\hat{p}_\rho)_{e^f e^r} | v^r \rangle \langle v^r | (\hat{p}_\rho)_{e^r e^g} | v^i \rangle}{\omega_{e^r v^r : e^g v^i} - \omega_1 - i\Gamma_r} \right.$$

$$\left. + \frac{\langle v^f | (\hat{p}_\rho)_{e^f e^r} | v^r \rangle \langle v^r | (\hat{p}_\rho)_{e^r e^g} | v^i \rangle}{\omega_{e^r v^r : e^f v^f} + \omega_1 + i\Gamma_r} \right\} \quad \textbf{2.20}$$

$$(\hat{p}_\rho)_{e^f e^r} = \langle e^f | \hat{p}_\rho | e^r \rangle \quad \textbf{2.21}$$

$$\omega_{e^r v^r : e^g v^i} = \omega_{e^r e^g} - \omega_{v^r v^i} \quad \textbf{2.22}$$

where e^f is the final electronic excited state, v^f is the final vibrational state, e^g is the ground electronic state, v^i is the initial vibrational state, e^r is the electronic eigenstate state r, and v^r is the vibrational eigenstate r.

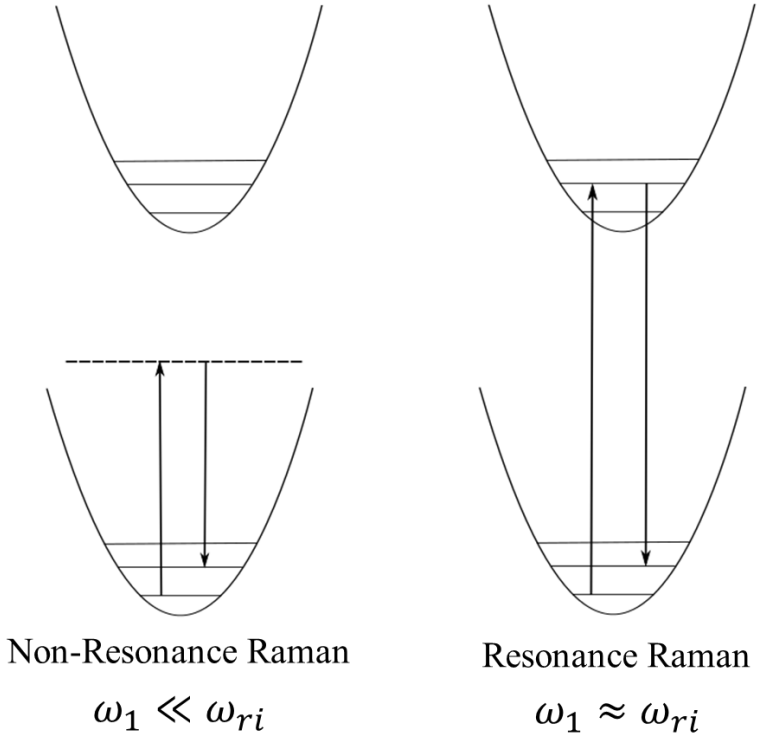

Figure 2.3: Difference between non-resonance Raman and resonance Raman scattering. In Raman scattering the energy of the incident light is far from any electronic transition. In resonance Raman the incident light energy is approximately equal to the energy required for an electronic transition.

From eq. 2.20 we can see how the resonance condition affects the polarizability and thus the intensity of Raman scattered light. As the incident light frequency (ω_1) approaches the absorption frequency $\omega_{e^r v^r : e^g v^i}$, the denominator of the first term in eq. 2.20 is minimized to $-i\Gamma_r$. As a result, the polarizability and Raman intensity is increased due to resonance.

For resonance Raman, equation 2.20 can be simplified. The Raman transition polarizability will be dominated by the resonant electronic transition and the contribution of non-resonant electronic transitions will be negligible in comparison. Thus, we can drop the sum over all electronic states r. In the resonance condition the second term of equation 2.20 is negligible compared to the first term and can be neglected. Finally, the Born-Oppenheimer principle allows us to separate the vibrational and electronic terms. Applying these simplifications produces:

$$(\alpha_{\rho\sigma})_{e^f v^f : e^g v^i} = \frac{1}{\hbar} (\hat{p}_\rho)_{e^f e^r} (\hat{p}_\rho)_{e^r e^g} \sum_{v^r} \left\{ \frac{\langle v^{f(g)} | v^{r(r)} \rangle \langle v^{r(r)} | v^{i(g)} \rangle}{\omega_{e^r v^r : e^g v^i} - \omega_1 - i\Gamma_r} \right\} \qquad \textbf{2.23}$$

For equation 2.23 to be non-zero the system must meet two conditions. Both transition dipole moments ($(\hat{p}_\rho)_{e^f e^r}$ and $(\hat{p}_\rho)_{e^r e^g}$) must be non-zero and both vibrational overlap integrals ($\langle v^f | v^r \rangle$ and $\langle v^r | v^i \rangle$) must be non-zero. If either condition is not met then the contribution of resonance to the transition polarizability and Raman intensity reduces to zero.

The condition for non-zero transition dipole moments requires the electronic transitions to be electric-dipole allowed. For the vibrational overlap integrals to be non-zero, the vibrational wave functions of states of v^f v^r and v^r v^i must not be orthogonal, ie. the Frank-Condon overlaps must be non-zero. For the situation where the vibrations in the two electronic states e^g and e^r are identical, the vibrational wave functions are orthogonal and the overlap integrals are

zero. For the vibrational overlap integrals to be non-zero the molecular vibration must have a different frequency in the excited state and/or there must be a shift in the potential energy minimum along the normal coordinate in the excited state (Figure 2.4).

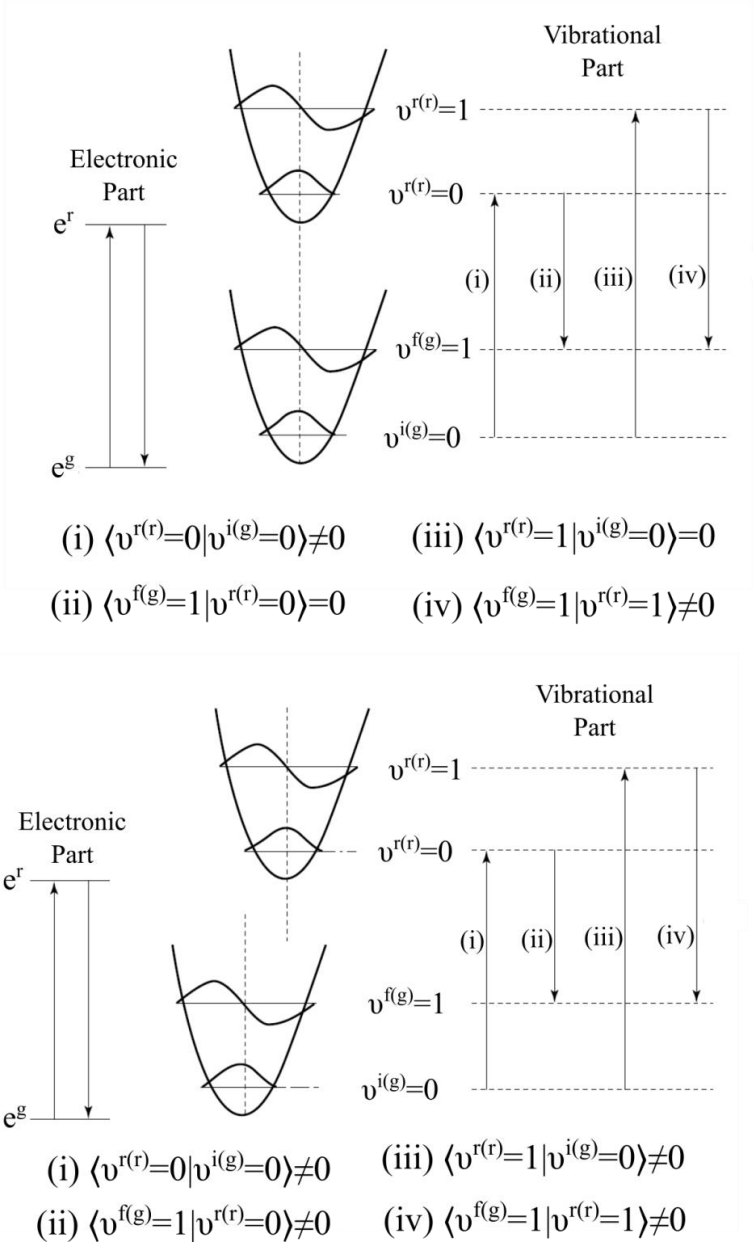

Figure 2.4: Potential energy curves and Frank-Condon overlaps for a vibrational motion without a shift (top) and with a shift (bottom) in the potential energy minimum along the vibrational normal coordinate. Adapted with permission by Long, D. A. The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules; John Wiley and Sons Ltd.: Chichester, 2002. Copyright 2002 John Wiley and Sons Ltd.

As a result, only vibrations with motions coupled to the resonant electronic transition are enhanced. This allows for the selective enhancement of a subset of a molecule's normal modes and is quite useful in examining the vibrational spectra of macromolecules such as proteins. Large molecules contain a large number of vibrations that result in spectral crowding. This makes examining individual spectral bands difficult. The selectivity of resonance Raman greatly reduces spectral crowding. In addition, by judiciously selecting the frequency of the incident radiation, one can choose to enhance vibrations that are coupled to a specific chromophore in molecules containing multiple chromophores. A great example of this is for the resonance Raman spectra of myoglobin with different excitation wavelengths (Figure 2.5). Incident light at ~200 nm is in resonance with the protein backbone $\pi \rightarrow \pi^*$ amide absorption band resulting in selective enhancement of vibrations of the protein backbone. In contrast, excitation at ~229 nm and ~415 nm is in resonance with tyrosine/tryptophan side chains and the heme group, respectively, enhancing vibrations that couple to these transitions. The frequencies and intensities of the protein's vibrations are sensitive to the structure and environment of the protein. Thus, as discussed in detail later, UVRR can be used to obtain detailed information on various chromophores within a protein.

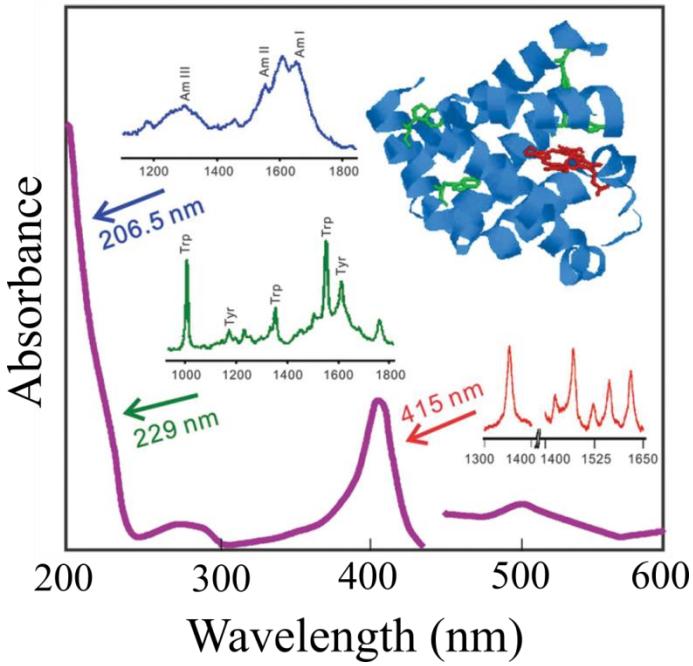

Figure 2.5: UVRR selective enhancement of normal modes in myoglobin. Incident excitation within different absorption bands of the molecule selectively enhance the Raman scattering for vibrations coupled to the electronic transition. Reprinted with permission from Oladepo et al., *Chem. Rev.*, 2012, 112(5), 2604-2628. Copyright 2012 American Chemical Society.

3.0 UV Resonance Raman of Proteins and Peptides

The selective enhancement of UVRR makes it a powerful tool for investigating protein structure and environment.[62,63] Careful selection of the incident light wavelength allows UVRR to probe vibrations found on different chromophores of a protein.[62] Certain UVRR band frequencies and intensities are dependent on the structure and environment of the resonant chromophore.[63] These spectroscopic markers can be used to obtain exquisite detail on protein structure. UVRR has extensively been used to probe the structure and solvation of the protein backbone[2,4–10] and various side chains.[46,66,70–72,82–85] In the following discussion, we will focus on how UVRR spectroscopy can be used to obtain information on the peptide backbone and Gln side chain, which is of importance to the UVRR study of polyQ tracts.

3.1 Peptide Backbone and Gln Side Chain Amide Group Chromophores

UVRR enhancement requires the incident light to be resonant with an absorption band of the analyte. Amide groups, such as the secondary amide of the peptide bond and primary amide of the Gln side chains, contain absorption bands in the deep UV (~180-190 nm) that can be probed with UVRR. Their strong absorption at ~180-190 nm results from the $\pi \rightarrow \pi^*$ electronic transitions of the primary and secondary amide groups. Raman excitation wavelengths within the amide absorption bands will selectively enhance vibrations of the amide group.

3.2 The UVRR Spectrum of Proteins

UVRR spectra of proteins collected with an excitation wavelength of ~200 nm are dominated by amide group Raman bands because the incident light is in resonance with the amide π→π* absorption band. This includes the secondary amide peptide bond and primary amide Gln side chains. Figure 3.1 shows a representative UVRR spectrum of a peptide excited with ~204 nm excitation.

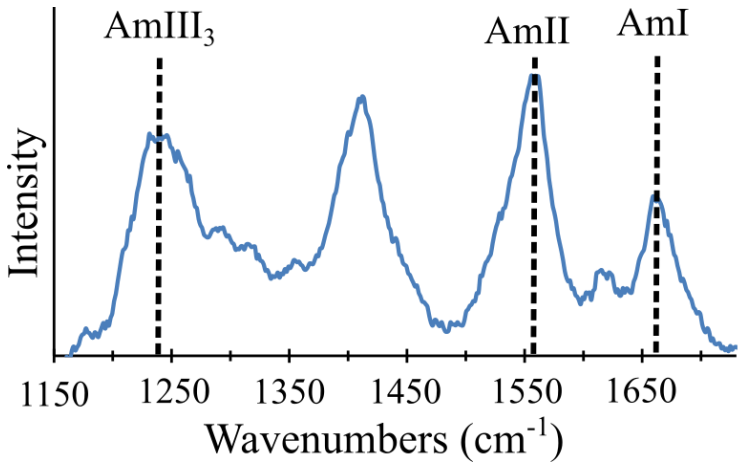

Figure 3.1: UVRR spectrum of a polyglutamine peptide ($D_2Q_{10}K_2$) excited with 204 nm excitation. Adapted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

Backbone amide π→π* excitation results in a large elongation of the C-N bond along with a smaller elongation of the C=O bond and contraction of the C-C_α and N-C_α bonds.[86] A

similar structural dependence upon excitation was previously found for primary amides.[72] Thus, the amide vibrations that have the most resonance enhancement generally contain C-N stretching motions. The dominant bands in the spectrum include the amide I (AmI), amide II (AmII), amide III (AmIII), and $C_\alpha H$ bands. The AmI band has a frequency of ~1600-1700 cm^{-1} and consists predominantly of C=O stretching motion[62] while the AmII band is found at ~1550 cm^{-1} and consists of out-of-phase N-H bending and C-N stretching. The AmIII bands are found in the ~1200-1400 cm^{-1} spectral region and consist predominately of in-phase N-H bending and C-N stretching.[62] The $C_\alpha H$ band is found at ~1350-1400 cm^{-1} and contains C_α-H bending and C-N stretching motions.[62] The motions of the amide vibrations are depicted in Figure 3.2. To avoid confusion we denote the amide bands of a primary amide with a "P" superscript (eg. AmI^P) and that of a secondary amide with an "S" superscript (eg. AmI^S).

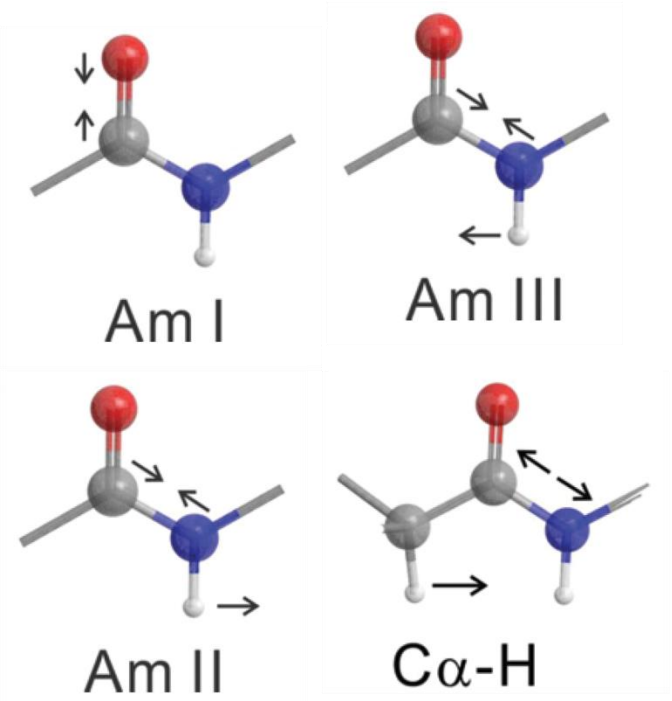

Figure 3.2: UVRR enhanced vibrational motions of amide groups. Adapted with permission from Oladepo et al., *Chem. Rev.*, 2012, 112(5), 2604-2628. Copyright 2012 American Chemical Society.

3.3 Separating Side Chain and Backbone Spectral Contributions

Because the secondary amide peptide backbone and primary amide Gln side chains have similar vibrational frequencies and UVRR enhancement, it is difficult to discern their UVRR bands. Xiong et al. previously developed a method to selectively highlight the spectral contributions of either the peptide backbone or Gln side chain by collecting UVRR spectra at different excitation wavelengths.[57]

Resonance Raman enhancement is proportional to the square of the molar absorptivity of the incident light.[87,88] The $\pi \rightarrow \pi^*$ absorption band of primary amides is at higher energy compared to that of secondary amides. Because of this, excitation deeper in the UV (~197 nm) increases the resonance enhancement for primary amide bands relative to secondary amide bands.

Xiong et al. showed that one can highlight the spectral contributions of the primary amide Gln side chains by exciting UVRR spectra of a protein at ~204 nm and ~197 nm and examining the 197-204 nm difference spectrum. In addition, the peptide backbone UVRR bands can be highlighted by subtracting the 197-204 nm difference spectrum from the 204 nm spectrum producing a 204-(197-204) nm difference spectrum. These techniques have been used extensively to differentiate the UVRR spectral contributions of a protein backbone and Gln side chains.[46,57,81,82]

4.0 UVRR Spectroscopic Markers Sensitive to Protein Structure and Environment

4.1 $AmIII_3$ Band of Peptide Backbone Monitors the Ψ Angle

Previously, Asher and co-workers discovered that the frequency of the $AmIII_3^S$ band of the peptide backbone sinusoidally depends on the Ramachandran Ψ angle.[65] The sinusoidal frequency dependence originates from coupling of the C_α-H bending vibration (found at ~1400 cm^{-1}) to the N-H bending motion of the of the $AmIII_3^S$ vibration. Coupling between the C_α-H and N-H groups is strongest when the peptide bond in a trans configuration resulting in a downshift of the $AmIII_3^P$ band. Thus, strong coupling and a downshifted $AmIII_3^S$ band occurs for extended strand structures such as PPII and β-strand structures while weaker coupling and an upshifted $AmIII_3^S$ band occurs for α-helical structures.

Mikhonin et al. characterized the effects of solution temperature and the peptide backbone solvation state on the $AmIII_3^S$ band frequency.[64] From this, they derived a set of equations that can be used to calculate Ψ angle from an experimental $AmIII_3^S$ frequency for a peptide backbone with various solvation environments and sample temperatures (Figure 4.1).

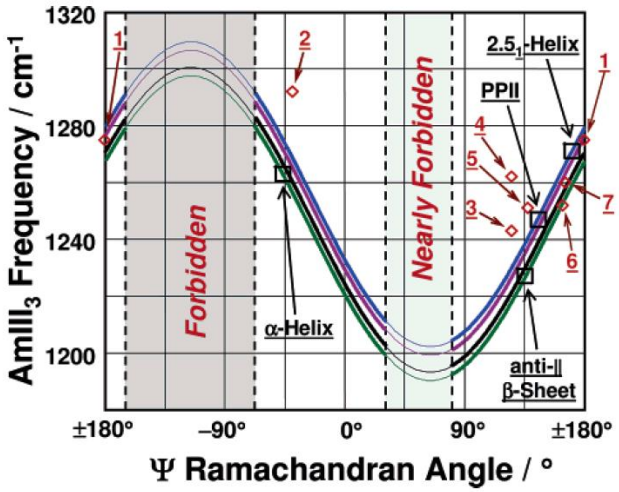

Figure 4.1: Sinusiodal dependence of the $AmIII_3^S$ band frequency on the Ψ Ramachandran angle of the peptide backbone. The curves show the correlation where the backbone is (blue) fully hydrated, (purple) only C=O groups are peptide-peptide hydrogen bonded, (black) only N-H groups are peptide-peptide hydrogen bonded, and (green) N-H and C=O groups are both peptide-peptide hydrogen bonded. The red diamonds show experimental data points for the crystalline peptides (1) Ala-Asp, (2) Gly-Ala-Leu, (3) Val-Glu, (4) Ala-Ser, (5) Val-Lys, (6) Ser-Ala, and (7) Ala-Ala. The black squares show experimental data points for peptides with various secondary structures. Reprinted with permission from Mikhonin et al. *J. Phys. Chem. B.* 2006, 110 (4), 1928–1943. Copyright 2006 American Chemical Society.

The $AmIII_3^S$ band is inhomogeneously broadened because a peptide contains a distribution of Ψ angles. Assuming that the inhomogeneous broadening is solely due to the Ψ angle distribution, Asher et al. showed that one can estimate the peptide Ψ angle distribution from the homogenous broadening of the $AmIII_3^S$ band.[74] First, the inhomogeneous $AmIII_3^S$ band is modeled as a sum of Lorentzian bands with a width equal to that of the homogenous $AmIII_3^S$ band. Raman shift is then converted into Ψ angle using one of the equations derived by Mikhonin et al.[64] The resulting Ψ angle probability distribution is plotted as a histogram.

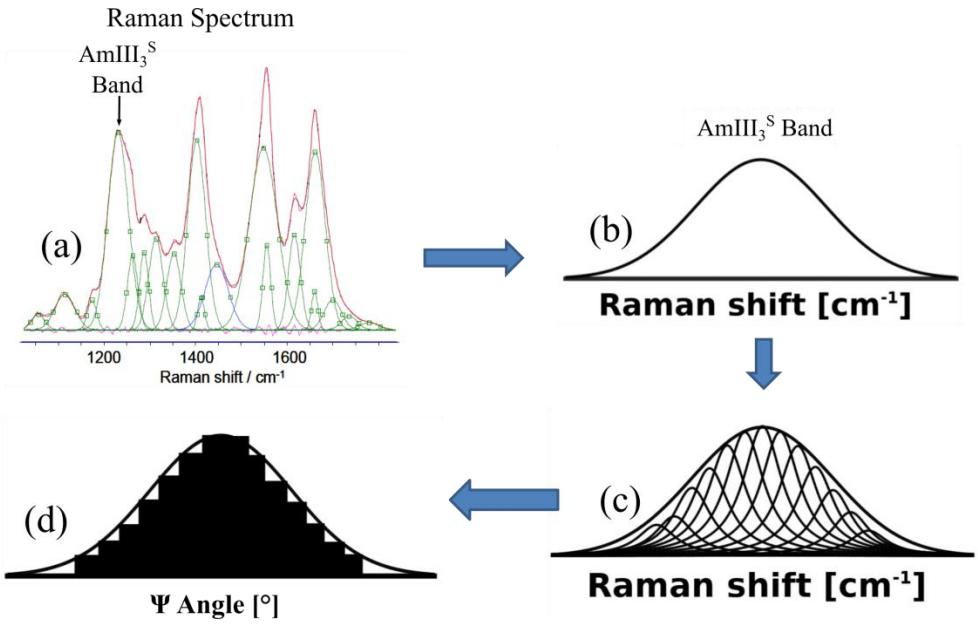

Figure 4.2: Calculating Ψ angle distribution from the inhomogenously broadened $AmIII_3^S$ band. (a) UVRR spectrum of peptide. (b) $AmIII_3^S$ band of peptide. (c) $AmIII_3^S$ band modeled as a sum of Lorentrzian bands with a breadth equal to the homogenous $AmIII_3^S$ band width. (d) Ψ angle distribution obtained by converting Lorentzian band frequency to Ψ angle.

Assuming that the Raman cross section is constant across the breadth of the $AmIII_3^S$ band, the intensity of the Lorentzian peak reports the relative population of peptide bonds with that Ψ angle. In addition, by applying the Boltzmann relation, one can estimate the Gibbs free energy landscape along the Ψ angle coordinate of a peptide. These methodologies have been extensively used to investigate the structure of proteins and peptides. [2,4–10]

4.2 $AmIII_3$ Band of Gln Side Chain Monitors the χ_3 Torsion Angle

Recently, Punihaole et al. discovered a UVRR marker band that reports on the Gln side chain χ_3 ($C_\beta C_\gamma$-$C_\delta O_\varepsilon$) torsion angle (Figure 4.3).[71] The structurally sensitive band is referred to as the $AmIII^P$ band because the vibration's motions are very similar to that of the $AmIII^S$ bands of the peptide backbone. The $AmIII^P$ band contains primarily C_β-C_γ stretching, $N_\varepsilon H_2$ rocking and C_δ-N_ε stretching motions of the Gln side chain.

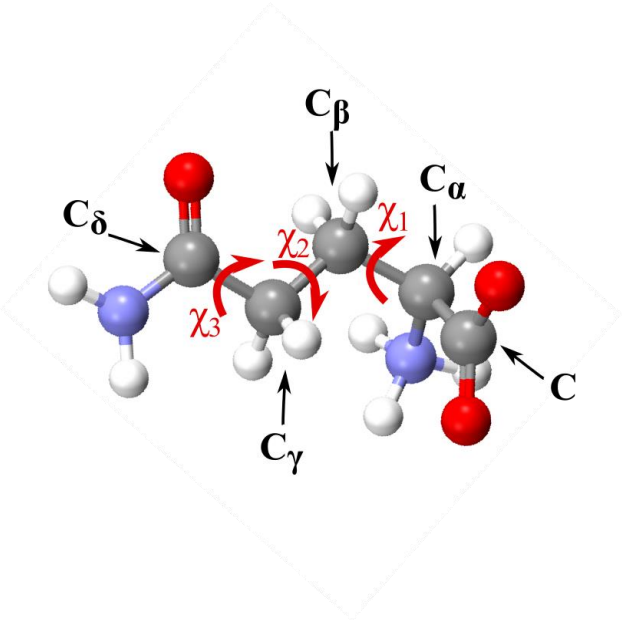

Figure 4.3: Structure and atomic labeling scheme of Gln. The χ_1, χ_2, and χ_3 torsion angles are labeled.

The AmIII^P band shows a sinusoidal frequency dependence on the Gln χ_3 torsion angle that originates from changes in hyperconjugation between the C_β-C_γ σ and C_δ=O_ε π* orbital as the χ_3 angle is rotated. At χ_3 angles approaching ~0° or ~180° the σ and π* orbitals do not overlap. In contrast, χ_3 angles approaching ±90° have significant σ/π* orbital overlap. As a result, electron density shifts from the σ to the π* orbital, which decreases the electron density and bond force constant of the C_β-C_γ bond. Thus, the AmIII^P band, which contains C_β-C_γ stretching motion, downshifts in frequency as χ_3 approaches ~±90°. Using a procedure similar to that described for the AmIII$_3^S$ band and Ψ angle correlation, one can estimate a χ_3 angle distribution and Gibbs free energy landscape from the inhomogeneously broadened AmIII^P band (See Section 4.1). This correlation was used to investigate the Gln side chain structure in small polyQ peptides in the solution[71] and fibril[46] state(s).

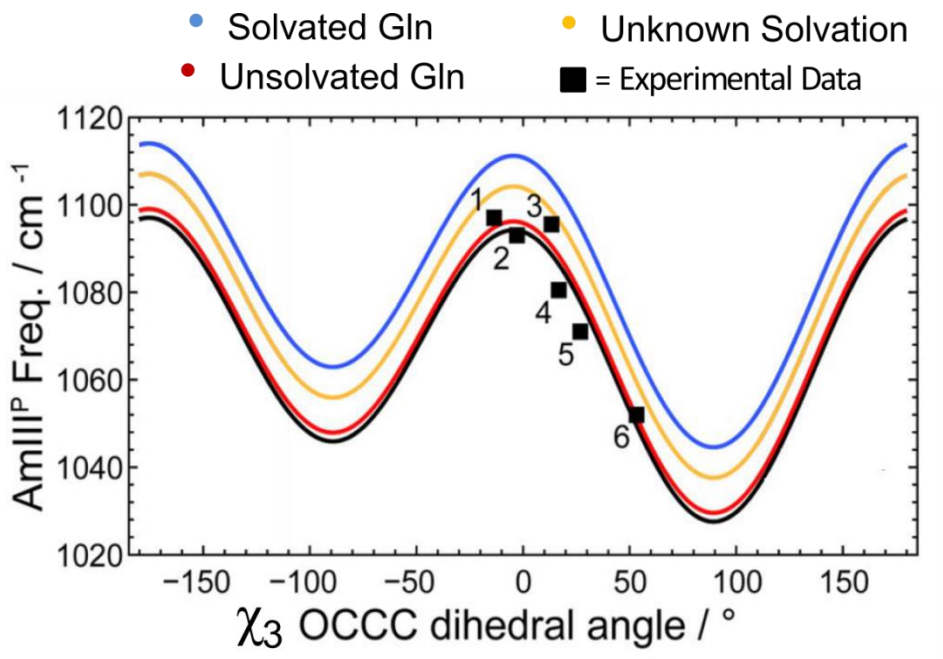

Figure 4.4: Dependence of the AmIII$_P$ frequency on the Gln side chain χ_3 torsion angle. The colored curves show the correlation for (black) crystalline experimental data, (blue) solvated Gln, (red) unsolvated Gln, and (yellow) Gln with unknown hydration. The black squares show experimental data points for crystalline (1) L-Gln, (2) Gly-Gln, (3) D-Gln, (4) Gln t-butyl ester, (5) N-Acetyl-Gln, and (6) Ser-Asn. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2015, 119 (41), 13039–13051. Copyright 2015 American Chemical Society.

4.3 AmI Band Sensitive to Hydrogen Bonding and Dielectric Environment

The AmI band of both primary and secondary amides was previously shown to depend on the hydrogen bonding and dielectric environment of the amide group. Various publications have shown that N-methylacteamide (NMA) (a model compound for the peptide backbone) has a solvation dependant AmI frequency and intensity.[89–93] Similarly, Punihaole et al. investigated the Raman spectrum solvation dependence of propanamide (a model compound for the Gln side chain) and found a dependence similar to that of NMA.[72]

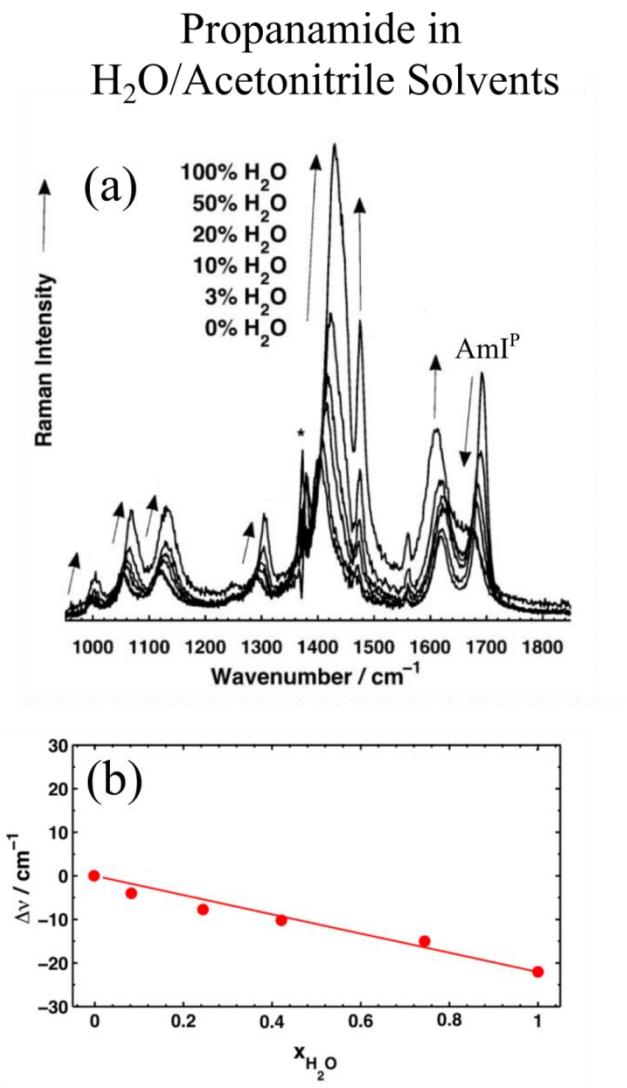

Figure 4.5: Dependence of the propanamide UVRR spectrum on solvation state. (a) UVRR spectra of propanamide in solvents containing different H_2O/Acetonitrile fractions. (b) Frequency dependence of the propanamide AmI band on water mole fraction. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2015, 119 (10), 3931–3939. Copyright 2015 American Chemical Society.

For both primary and secondary amides, the AmI band frequencies and intensities were found to decrease with increasing dielectric and hydrogen bonding environment of the solvent. These trends result from changes in the structure of the amide group. In a high dielectric environment, the amide group polar resonance structure ($^{-}$O-C=NH^{+}) increases in stability compared to the nonpolar structure (O=C-NH). Because the C=O bond force constant is weaker in the polar resonance structure, the AmI C=O stretching frequency decreases as the dielectric environment increases. In addition, the formation of stronger hydrogen bonds between the carbonyl (C=O) group and a hydrogen bond donor results in a decrease in the carbonyl force and a downshifted the AmI frequency.

These structural changes also affect the AmI resonance Raman intensity. This occurs because changes between the ground and excited state normal coordinate displacement alters the Frank-Condon overlaps of the resonance Raman transition (see section 2.3). For example, the resonance enhancement of a totally symmetric normal mode is approximately proportional to the square of the displacement between the ground and excited state equilibrium geometries along the enhanced vibrational normal coordinate.[94,95] The C-N and C=O bonds of the amide group are elongated in the $\pi \rightarrow \pi^*$ excited state compared to the ground state.[96] In a higher dielectric environment or in a stronger hydrogen bonding environment, the ground state C=O bond is elongated. This decreases the C=O bond displacement between the ground and excited state, resulting in less AmI band resonance enhancement.

5.0 Monomeric Polyglutamine Structures that Evolve into Fibrils

Adapted with permission from: David Punihaole, Ryan S. Jakubek, Riley J. Workman, Lauren E. Marbella, Patricia Campbell, Jeffry D. Madura, and Sanford A. Asher. Monomeric polyglutamine structures that evolve into fibrils. The Journal of Physical Chemistry B, 2017, 121(24), pp.5953-5967.

Author Contributions: D.P. and R.S.J. contributed equally to this work. D.P and R.S.J. collected, analyzed, and interpreted UVRR data. Computational work was performed by R.J.W. and analyzed by R.J.W with assistance by D.P., J.D.M., and R.S.J. DOSY NMR was performed by L.E.M and TEM images were collected by P.C. The manuscript was prepared by R.S.J., D.P., and S.A.A. with assistance by R.J.W.

We investigate the solution and fibril conformations and structural transitions of the polyglutamine (polyQ) peptide, $D_2Q_{10}K_2$ (Q10), by synergistically using UV Resonance Raman (UVRR) spectroscopy and Molecular Dynamics (MD) simulations. We show that Q10 adopts two distinct, monomeric solution conformational states, a collapsed β-strand and a PPII-like structure that do not readily interconvert. This clearly indicates a high activation barrier in solution that prevents equilibration between these structures. Using metadynamics, we explore the conformational energy landscape of Q10 to investigate the physical origins of this high activation barrier. We develop new insights into the conformations and hydrogen bonding environments of the glutamine side chains in the PPII and β-strand-like conformations in solution. We also use the secondary structure-inducing cosolvent, acetonitrile, to investigate the conformations present in low dielectric constant solutions with decreased solvent-peptide hydrogen bonding. As the mole fraction of acetonitrile increases, Q10 converts from PPII-like

structures into α-helix-like structures and β-sheet aggregates. Electron microscopy indicates that the aggregates prepared from these acetonitrile-rich solutions show morphologies similar to our previously observed polyQ fibrils. These aggregates re-dissolve upon the addition of water! These are the first examples of reversible fibril formation. Our monomeric Q10 peptides clearly sample broad regions of their available conformational energy landscape. The work here develops molecular-level insight into monomeric Q10 conformations and investigates the activation barriers between different monomer states and their evolution into fibrils.

5.1 Introduction

Expansions of genomic "CAG" codon repeats are associated with at least 10 neurodegenerative disorders, including Huntington's disease.[1] CAG repeats encode for expanded polyglutamine (polyQ) tracts in proteins. These expanded polyQ tracts cause proteins to aggregate into amyloid-like fibrils, which are the pathological hallmarks of CAG repeat diseases. The penetrance and severity of these diseases correlate with the length of the polyQ repeat expansion. In Huntington's disease,[5,97] for example, repeats between 17–30 glutamines in the huntingtin protein are generally considered benign, whereas repeat lengths that exceed 36 residues typically result in fibrils and the manifestation of disease symptoms.

Although flanking sequences are known to influence aggregation kinetics and mechanisms,[40,48] the expanded polyQ protein segments are the only apparent commonality shared by CAG repeat diseases. For this reason, investigating the structures of polyQ peptides enables the molecular understanding of how glutamine repeats cause protein misfolding and

aggregation. Indeed, previous studies show that aggregates prepared *in vitro* from model polyQ peptides share many characteristic features with pathologically relevant amyloid-like fibrils, including displaying filamentous morphologies, binding Thioflavin-T, and exhibiting β-sheet-rich structures.[35,36]

Many experimental studies suggest that soluble polyQ peptides are structurally disordered in aqueous solution.[34–36,41,98] Other studies, however, suggest that polyQ peptides contain small populations of "folded" structures, which are hypothesized to be putative cytotoxic agents.[16] Computational investigations[37–39,99] support the experimental finding that polyQ peptides, regardless of repeat length, are largely disordered, with only transient elements of regular secondary structures such as α-helices, β-sheets, and turns.

Although polyQ peptides are structurally disordered, they do not behave as true random coil polymers.[97] Several studies[37,38,43–45,61,100] suggest that the end-to-end distances of polyQ peptides deviate significantly from random coils, since they adopt relatively collapsed, globule-like structures. In addition, detailed analyses of circular dichroism (CD) and nuclear magnetic resonance (NMR) spectra[98] show that that polyQ peptides possess a high propensity towards adopting polyproline II-like (PPII-like) secondary structures, although not necessarily forming long, continuous segments.

Despite numerous studies, several important aspects of polyQ peptide structure in solution remain poorly understood. For example, little is known about the solution-state structural change(s) of polyQ peptides that lead to fibril nucleation and growth. One model, by Wetzel and coworkers,[35] proposes that fibril nucleation is initiated by an energetically unfavorable structural conversion from "random" coil to β-sheet. An alternative model, proposed by Pappu and coworkers,[38] argues that polyQ peptides adopt disordered globule structures, which

non-specifically aggregate into high molecular weight oligomers that subsequently convert into β-sheet-rich fibrils.

A detailed understanding of the conformational energy landscape that controls the solution-state structural propensities of polyQ peptides would yield valuable insights that could resolve some of the current debates in the field. Unfortunately, standard biophysical methods only achieve limited insight into the solution state structure of disordered proteins and peptides, such as polyQ.[101,102] Incisive progress requires the use of new experimental and computational biophysical tools.

In the work here, we investigate the ensemble of solution-state structures and the conformational energy landscape of the model peptide system, $D_2Q_{10}K_2$ (Q10), by synergistically utilizing UVRR spectroscopy and MD simulations. This work builds upon our previous study by Xiong et al.[57]

Figure 5.1 schematically summarizes the new insights that we develop in this work regarding the conformational energy landscape of Q10. We find that Q10 can be poised to exist in two distinct, *monomeric* conformational states in aqueous solution. One of these states is in a collapsed β-strand-like structure (Figure 5.1a) that readily aggregates into amyloid-like fibrils (Figure 5.1e). The other state is a predominately PPII-like structure (Figure 5.1b) that resists aggregation.

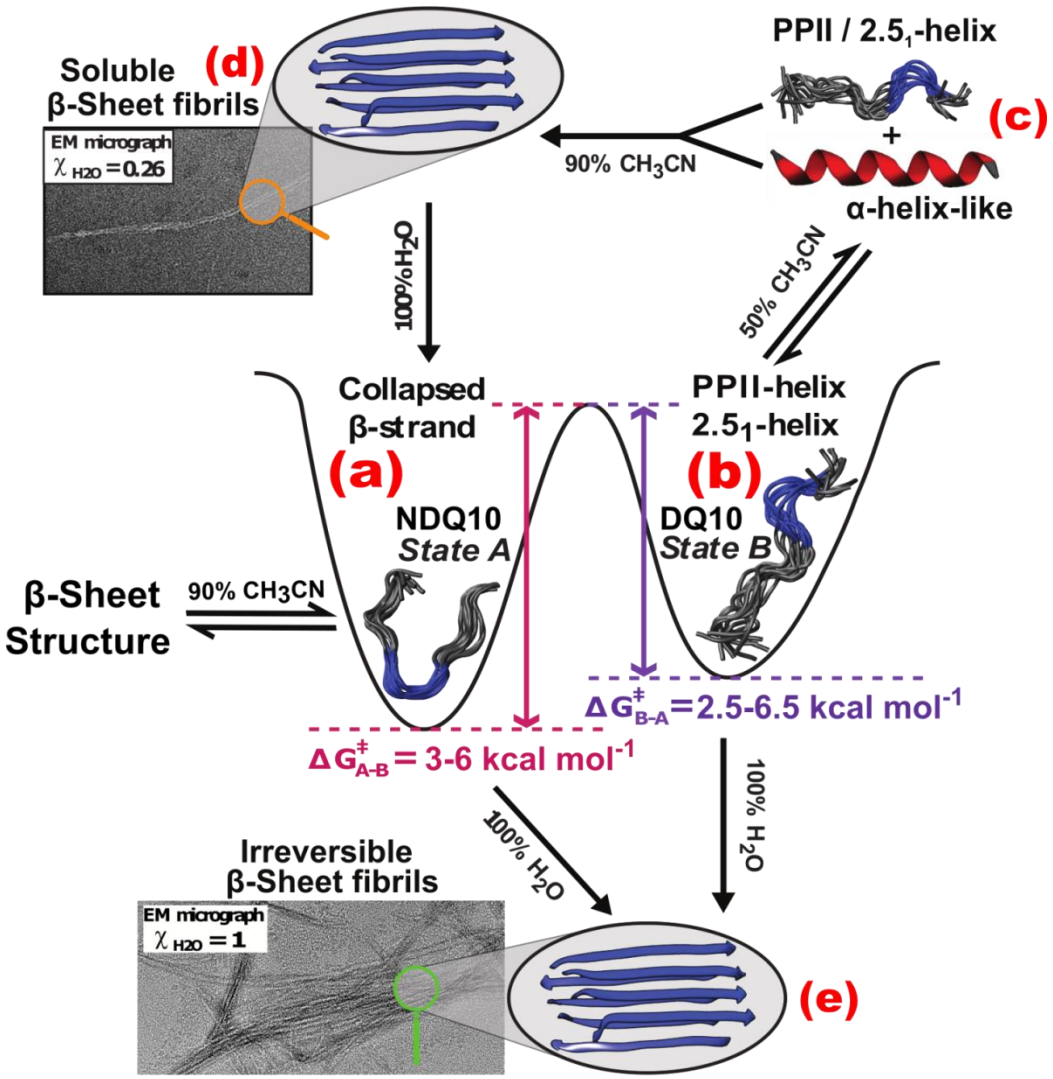

Figure 5.1: Summary of polyQ solution energy landscapes.

Figures (a) and (b) show the potential wells for the aqueous Q10 β-strand and PPII states observed. Conformation (c) depicts the α-helix-like conformations observed in ~50% acetonitrile/water solutions. Conformation (d) derives from Q10 β-sheet fibrils formed in high acetonitrile concentrations. These fibrils re-dissolve in high water content acetonitrile/water solutions. Conformation (e) derives from irreversibly formed Q10 β-sheet fibrils grown in water. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

The collapsed β-strand and PPII-like structures do not readily interconvert, which indicates that there is a large energy barrier that prevents these two states from equilibrating in solution. Using metadynamics, we simulate the conformational energy landscape of Q10 and investigate the origin of this high energy barrier. Overall, we find excellent, *quantitative* agreement between the simulation results and our UVRR data. This enables us to use our experimentally validated simulations to gain detailed, molecular-level structural insights into the conformational ensemble of Q10 in solution.

We also investigate the impact of low hydrogen bonding and dielectric environments on Q10 peptide structure. We measure the UVRR spectra of Q10 in different mixtures of acetonitrile and water. We find that, at moderate concentrations of acetonitrile, PPII-like conformations convert into structures comprised of predominately α-helix-like and turn structures (Figure 5.1c), while high concentrations of acetonitrile induce the formation of β-sheet fibrils (Figure 5.1d) and non-fibrillar aggregates. Surprisingly, fibrils prepared from high acetonitrile concentrations re-dissolve in water-rich solutions.

5.2 Experimental Section

5.2.1 Materials

The 14-residue $D_2Q_{10}K_2$ (Q10) peptide was purchased as a lyophilized powder from CPC Scientific at ≥95% purity. Trifluoroacetic acid (TFA) and 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) were purchased at ≥99% purity from Acros and Fluka, respectively. Far-UV grade

acetonitrile (CH_3CN) was purchased from Acros. Optima-grade H_2O was purchased from Fisher Scientific, and D_2O (99.9 atom % D) was purchased from Cambridge Isotope Laboratories, Inc.

5.2.2 Sample Preparation

Non-disaggregated Q10 (NDQ10) and disaggregated Q10 (DQ10) samples were prepared as described previously by Xiong et al.[57] NDQ10 solutions were prepared by directly dissolving the peptide in pure water. DQ10 solutions were prepared by utilizing a disaggregation protocol based on the one developed by Wetzel and coworkers.[31] Briefly, DQ10 samples were prepared by suspending the lyophilized peptide powder in a 1:1 (v/v) mixture of TFA and HFIP. The samples were sonicated for 20 min and incubated at room temperature for 2 h. The solvents were evaporated under a gentle stream of dry N_2 gas for ~30 min. The peptide film was dissolved in 1 mL of water. DQ10 peptide solutions were then diluted to their final concentrations with water. Raising the pH of monomeric NDQ10 increases its propensity to aggregate. To avoid this, we investigated both NDQ10 and DQ10 solutions at pH values that ranged between pH 2 to pH 3. The NDQ10 monomer solutions were also found to be very sensitive to impurities that nucleated aggregation into light scattering particles, which we presumed were dominated by β-sheet conformations. We scrupulously cleaned all vials used for sample preparation to remove nucleating centers.

5.2.3 UV Resonance Raman (UVRR) Spectroscopy

The UVRR instrumentation used is described in detail by Bykov et al.[103] The ~204 nm UV light was generated by Raman shifting the third harmonic of a Nd:YAG Infinity laser (Coherent, Inc.) in H_2 gas (~30 psi) and selecting the fifth anti-Stokes line. The laser light was focused onto a spinning Suprasil quartz NMR tube containing the samples. A ~165° backscattering geometry was used. The scattered light was dispersed using a home-built subtractive double monochromator and detected with a liquid N_2 cooled, back-thinned Spec10:400B CCD camera (Princeton Instruments) with a Lumogen E coating. The spectrometer resolution was ~5 cm^{-1}. A detailed description of the spectral processing and fitting is given in Appendix A.

5.2.4 NMR Diffusion Measurements

All ^1H diffusion NMR spectra were acquired with an 11.7 T magnet with a Bruker AVANCE III 500 console (Bruker Biopsin, Billerica, MA) at 298 K. NDQ10 and DQ10 solutions were prepared at 1 $mg \cdot mL^{-1}$ in 90% D_2O. Spectra were recorded with a stimulated echo bipolar pulsed field gradient pulse sequence. The WATERGATE pulse sequence was used for water suppression.[104] The maximum strength of the gradient coil was 0.5 $T \cdot m^{-1}$ and was varied from 5–95% during the measurements. The decay of the NMR signal integrated area, *I*, as a function of the applied gradient strength, *G*, is described by the Stejskal-Tanner equation:[105]

$$I = I_0 e^{-[(\gamma G\delta)^2(\Delta - \frac{\tau}{2} - \frac{\delta}{8})D]} \quad \textbf{5.1}$$

where I_0 is the initial intensity, γ is the gyromagnetic ratio of ^{1}H, δ is the length of the gradient pulse (2.5 ms), Δ is the diffusion time (50 ms), τ is the time between bipolar gradient pulses, and D is the apparent diffusion coefficient.

5.2.5 Transmission Electron Microscopy (TEM)

Quantifoil 400 mesh copper grids were glow discharged using an EmiTech Glow Discharge unit for 70 s at 25 mA. Aliquots of 4 μL Q10 solutions were placed onto the freshly prepared glow discharged grids, rinsed with double distilled H_2O, and stained with 4% (w/v) uranyl acetate. Samples were imaged in a FEI Tecnai 20F Electron Microscope. Images were collected using a Gatan Ultrascan 4K CCD camera at 30,000 and 50,000× magnification.

5.3 Computational Methods

5.3.1 General Simulation Details

Metadynamics simulations were used to explore the Q10 conformational landscape. Constrained MD simulations were used to investigate the stabilities of the 2.5_1-helix and PPII structures. The different peptide conformations were all constructed with the Molecular

Operating Environment (MOE 2013.10) software suite.[106] The initial, solvated Q10 peptides were energy minimized for 10000 steps using the conjugate gradient method. All peptide models were solvated with 6681 TIP3P water molecules in a periodic box with dimensions of 60 × 60 × 60 A^3. The NAMD software package (version 2.11)[107] was used for all MD simulations. The trajectory data were analyzed using VMD 1.9.2 and native Tcl scripting.[108] All MD simulations were run at the TACC supercomputer facility, while Q10 peptide models were prepared and energy minimized on a 4-core iMac desktop.

Potential energies were calculated with the CHARMM36 force field.[109] CHARMM36 contains the CMAP[110] dihedral corrections intended to decrease α-helix bias and stabilize β-strand secondary structure. The particle mesh Ewald algorithm[111] was used with a grid spacing of 1.0 A to calculate full system electrostatics. An integration time step of 2 fs was used. All simulations were carried out under NPT conditions, with a Langevin thermostat and piston to regulate the temperature at 300 K and pressure at 1.01325 bar.[112] The pair list distance was 14.0 Å, the pair interaction cutoff was 12 .0 Å, and the switch distance was 10.0 Å.

5.3.2 Metadynamics Simulation Details

The initial Q10 monomer used in the metadynamics simulation was constructed by using a fully extended peptide with Ramachandran (Φ, Ψ) dihedral angles of 180°. The initial, solvated Q10 peptide systems were energy minimized for 10000 steps using the conjugate gradient method followed by 50 ps of equilibration. The data used to construct the free energy landscape was collected over the next 1.0 μs of simulation time.

To specify the Q10 conformational space, we defined three collective variables that monitored the root mean square deviation (RMSD) of a Q10 structure relative to the idealized α-helix, β-hairpin, and PPII reference structures. The α-helix and PPII reference structures were created with MOE, whereas the β-hairpin reference structure was taken from preliminary metadynamics simulation results (coordinate files available for download). The RMSD was measured using only the peptide α-carbon positions. The maximum RMSD coordinate was set to 12.0 Å from the reference structure. This upper limit was maintained by applying a 1.0 kcal · mol^{-1} · $Å^{-1}$ half-harmonic potential constraint. This three dimensional conformational energy landscape approach has been successfully employed by us previously.[113–115]

In our metadynamics simulations, artificial potentials in the form of Gaussian functions were added every 500 steps. The height of added Gaussian functions started at 1.0 kcal · mol^{-1}, and their width was 3.0 Å. Well-tempered metadynamics[116] was used to streamline sampling of the conformational energy landscape. This protocol is a variation of the original metadynamics algorithm developed by Laio and Parrinello[117] that utilized a gradual decrease in the height of added Gaussians throughout the simulation. This allowed shallow energy wells to be filled in quickly, while deeper energy wells filled more slowly. This method allows the deeper energy wells to be explored in finer detail during the simulation data collection run.

Coordinates for low-energy structures were extracted from the trajectory based on RMSD values and the Gaussian deposition history. Hydrogen bonding analysis of structures found in low energy wells was performed with VMD's Hydrogen Bond analysis module. The Ψ and χ_3 dihedral angles were extracted from the simulation trajectories with a Tcl script, which is available for download. A heavy atom distance of <3.5 Å and a bond angle of 18030) was used to define a hydrogen bond. Metadynamics free energy data and collective variable history were

processed with two respective Python scripts, which are available for download. Script 1 identifies the energy landscape positions for structures of interest from the potential of mean force (pmf) file. Script 2 searches the metadynamics trajectory and identifies the structures that match the RMSD values found using script 1.

5.3.3 Constrained MD Simulation Details

We used constrained MD simulations to investigate the energies of Q10 in PPII and 2.5_1-helix conformations constructed with (Φ, Ψ) angles of (−75°, 150°) and (−130°, 177°), respectively. In the constrained MD simulations, harmonic constraints were placed on the Φ and Ψ dihedral angles for the Q10 peptides. These constraints ensured that the 2.5_1-helix and PPII peptides would retain their respective secondary structures throughout the simulations. A force constant of 0.25 kcal · mol^{-1} · $Å^{-1}$ was used for each dihedral constraint and the constraints were applied with the Collective Variables module in NAMD.

For the constrained MD simulations the solvated Q10 peptides were energy minimized for 10000 steps using the conjugate gradient method. The peptide systems were equilibrated for 1 ns, followed by data collection runs of 200 ns. Peptide potential energies were calculated using the NAMD Energy module. Bond, angles, dihedral angles, van der Waals interaction energies, and electrostatic energies were calculated for every frame of the trajectory output files. Energies were averaged using a simple Python script. The hydrogen bonding analysis of these constrained MD simulations was also performed with VMD's Hydrogen Bond module. The hydrogen bond parameters were the same as those described in the metadynamics section above.

5.4 Results And Discussion

To maximize clarity, we include numbered bold headings that highlight the most important polyQ structural discoveries made here that are visually summarized in Figure 5.1.

5.4.1 Solution-State Structures of Q10 in H_2O

Our previous study by Xiong et al.[57] examined the solution-state conformations of Q10, and suggested that it exists in two different solution-state conformations that depend on the sample preparation method. Preparing Q10 using a standard "disaggregation" protocol[31] (forms a solution containing the peptide denoted as DQ10, see Experimental Section for details) results in an ensemble of predominately PPII-like peptide conformations. In contrast, non-disaggregated Q10 (denoted NDQ10) dissolved in water adopts a distinctly different conformation that Xiong et al.[57] proposed was a mainly β-hairpin conformation.

We re-measured the UVRR spectra of DQ10 and NDQ10. Figure 5.2 shows representative ~204 nm excited UVRR spectra of the DQ10 and NDQ10 peptides dissolved in H_2O. The DQ10 UVRR spectra are essentially identical to those measured previously by Xiong et al.,[57] who investigated DQ10 peptide solutions at pH 7. In contrast, the NDQ10 UVRR spectra (Figure A.7) differ significantly from that of Xiong et al.[57] It appears that Xiong et al's.[57] NDQ10 samples had begun to aggregate. The sensitivity of NDQ10 monomers to nucleation of aggregation may have caused these spectral differences. Thus, we have altered our assignment of the NDQ10 solution monomer conformation.

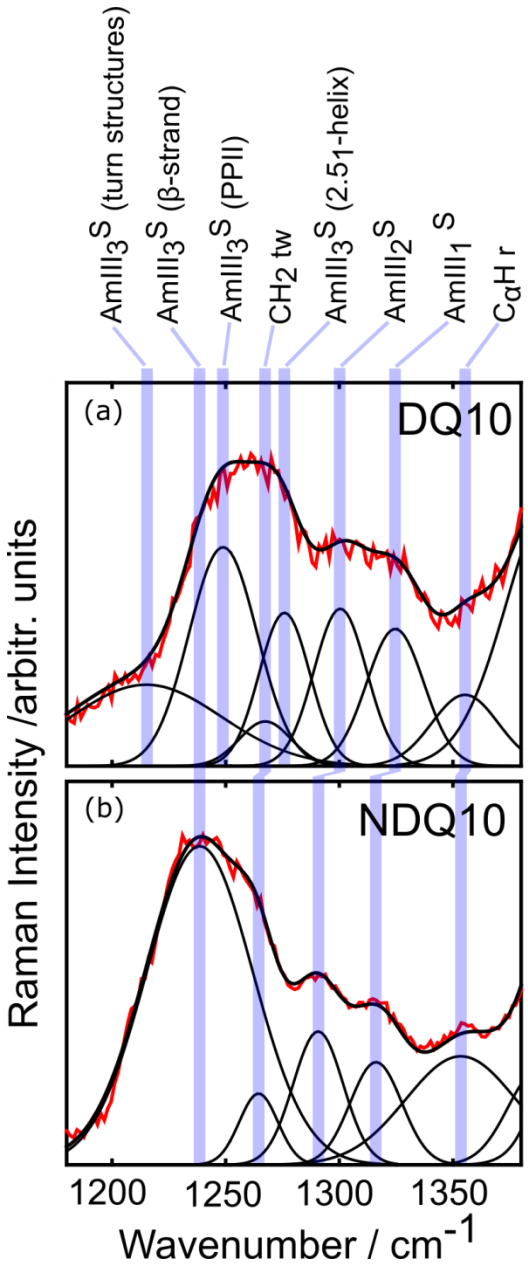

Figure 5.2: Representative 204 nm UVRR spectra of NDQ10 and DQ10.

Assignments of the AmIII^S region for (a) DQ10 and (b) NDQ10 are shown. The peptide solutions were prepared at 0.3 mg·mL^{-1} concentrations. Spectra from three independent replicates were globally fit using a minimum sum of Gaussian and Lorentzian bands, as described in Appendix A. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

5.4.2 UVRR Shows that DQ10 Populates PPII and 2.5_1-helix Conformations, While NDQ10 Populates β-strand-like Conformations

UVRR excitation at ~200 nm occurs within the NV_1 electronic transitions of the secondary and primary amides of the peptide backbone and glutamine side chains[57]. Thus, the Figure 5.2 ~204 nm excitation UVRR spectra of polyQ peptides are dominated mainly by resonance enhanced bands that derive from primary and secondary amide (Am) vibrations, which we denote with the superscripts ^{P} and ^{S}, respectively.

The Figure 5.2 spectra show the structurally sensitive extended $AmIII^S$ region between 1180 to 1380 cm^{-1}.[62] The DQ10 and NDQ10 spectra were curve-fit using a new, more rigorous and self-consistent global analysis, as described in Appendix A (Figure A.1). Both the DQ10 and NDQ10 UVRR spectra show bands at ~1355 cm^{-1}, ~1320 cm^{-1}, and ~1295 cm^{-1}. The 1355 cm^{-1} features are assigned to a mainly C_α–H rocking mode based on our previous normal mode analyses of glutamine.[41] The peaks at 1320 cm^{-1} and 1295 cm^{-1} are the $AmIII^S_1$ and $AmIII^S_2$ bands, respectively[62,64,65,118,119]. However, we note that the $AmIII^S_1$ and $AmIII^S_2$ overlap with additional weaker bands that are due to CH_2 twisting modes of the glutamine side chains.

The $AmIII^S_3$ region of DQ10 (Figure 5.2a) is similar to that of Xiong et al. with bands at ~1275 cm^{-1}, ~1265 cm^{-1}, ~1250 cm^{-1}, and ~1215 cm^{-1}. In contrast the $AmIII^S_3$ region of NDQ10 contains bands at ~1240 cm^{-1} and ~1265 cm^{-1}, which significantly differs from spectra reported by Xiong et al.[57] The difference in the NDQ10 solution state spectra is most likely due to some aggregation of Xiong et al.'s NDQ10 sample, since their spectrum is more similar to that of our NDQ10 fibrils.[46] Additionally, our recent insights into the UVRR spectra of glutamine shows that the ~1265 cm^{-1} band derives from CH_2 wagging of the glutamine side chain instead of

an $AmIII^{S}_{3}$ vibration.[71] As discussed below, our new understanding of the UVRR spectrum of NDQ10 modifies our assignment of the NDQ10 solution conformation.

5.4.3 Ramachandran Ψ Angle Distributions of DQ10 and NDQ10

The $AmIII^{S}_{3}$ band is the most structurally sensitive spectroscopic marker of the peptide backbone conformation because its frequency depends on the Ramachandran Ψ dihedral angle.[64,65] As discussed in detail by Asher and coworkers,[65] this sinusoidal frequency dependence derives from the coupling of the C_{α}–H bending vibration with the N–H bending component of the $AmIII^{S}_{3}$. This coupling is strong for β-strand and PPII-like structures, where the C_{α}–H and N–H groups are approximately in a cis-configuration, which downshifts the $AmIII^{S}_{3}$ frequency. In contrast, for α-helical structures, the C_{α}–H and N–H groups are in a trans configuration. This decouples the C_{α}–H and N–H bending motions, which results in an upshift of the $AmIII^{S}_{3}$ band frequency.

We utilized the structural sensitivity of the $AmIII^{S}_{3}$ band to estimate the Ramachandran Ψ angle distributions for DQ10 and NDQ10 in solution. To do this, we employed the methodology of Mikhonin et al.[64] (see Appendix A for details), which correlates the frequencies of the $AmIII^{S}_{3}$ band envelopes to the peptide bond Ψ angles. This enables us to roughly estimate the probability distribution of peptide bond Ψ angles that derive from the inhomogenously broadened $AmIII^{S}_{3}$ bandshape.

Figure 5.3a and c shows the Ψ angle distributions of DQ10 and NDQ10. The Ψ angle distribution of DQ10 is essentially identical to that reported by Xiong et al.,[57] consisting of three

prominent peaks located at ~10°, ~150°, and ~175°. The peak located at 150° is indicative of PPII-like conformations, while the peak centered at 10° indicates the presence of turn-like structures. The peak centered at 175° suggests that a significant fraction of DQ10 peptide bonds adopt 2.5_1-helix-like conformations.

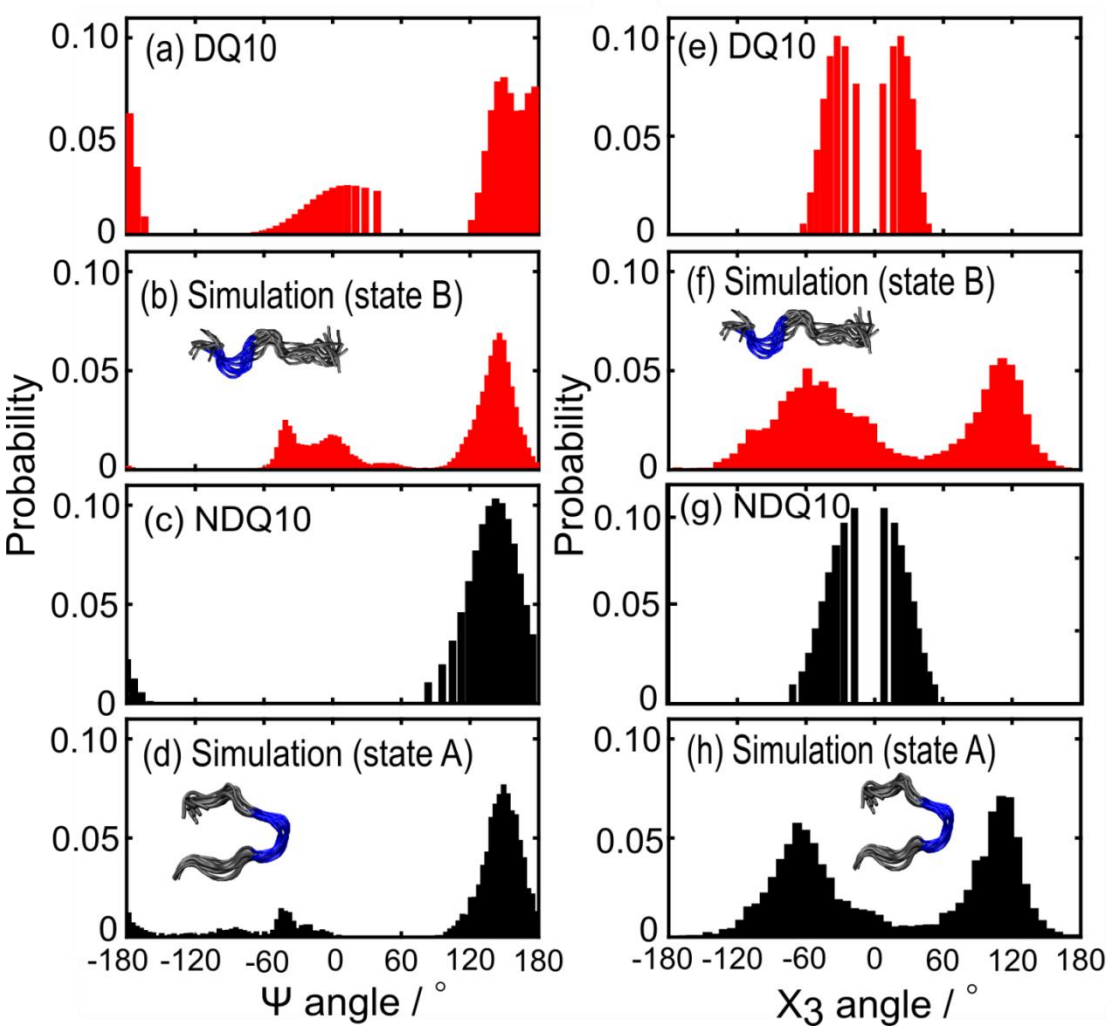

Figure 5.3: Comparison of Ψ and χ_3 angle distributions measured by UVRR and calculated by MD. Comparison between DQ10 Ψ angle distributions measured by (a) UVRR to (b) MD structures found in state B. Comparison between NDQ10 Ψ angle distributions measured by (c) UVRR to (d) MD structures found in state A. Comparison between DQ10 χ_3 angle distributions measured by (e) UVRR to (f) MD structures found in state B. Comparison between NDQ10 χ_3 angle distributions measured by (g) UVRR versus (h) MD structures found in state A. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

The NDQ10 Ψ angle distribution (Figure 5.3c) shows a peak located at ∼140° indicative of *β*-strand conformations. *In contrast to the NDQ10 Ψ angle distribution of Xiong et al., we do not observe a peak at Ψ ∼−40° that they assigned to turn structures.*

5.4.4 Hydrogen Bonding Environment of DQ10 and NDQ10 Side Chains

Our previous study of primary amide vibrations shows that the AmI^P and $AmII^P$ bands report on the hydrogen bonding and dielectric environments of the glutamine side chain C=O and NH_2 groups, respectively.[90] Xiong et al. showed that ∼198 nm excitation enhances the primary amide UVRR bands significantly more than the secondary amide bands.[57] As a result, the primary amide bands can be selectively studied by calculating the difference spectrum between ∼198 nm and ∼204 nm excitation UVRR spectra.

Figure 5.4 shows the 198 nm – 204 nm UVRR difference spectra of DQ10 and NDQ10. The AmI^P (mostly C=O stretching) and $AmII^P$ (NH_2 scissoring) bands of DQ10 are located at ∼1680 cm^{-1} and ∼1614 cm^{-1}, respectively. In NDQ10, the AmI^P frequency downshifts to ∼1660 cm^{-1}, although a prominent shoulder remains at ∼1680 cm^{-1}. As with DQ10, the $AmII^P$ band of NDQ10 is located at ∼1614 cm^{-1}. The most intense bands in the 198 nm – 204 nm UVRR difference spectra appear at ∼1437 cm^{-1} for DQ10 and ∼1414 cm^{-1} for NDQ10. We assign these bands to a complex vibration that contains CH_2 wagging, as well as significant contributions of C–C stretching, CH_2 scissoring, and side chain C-N stretching. The C-N stretching component of this vibration likely contributes to its UVRR enhancement.

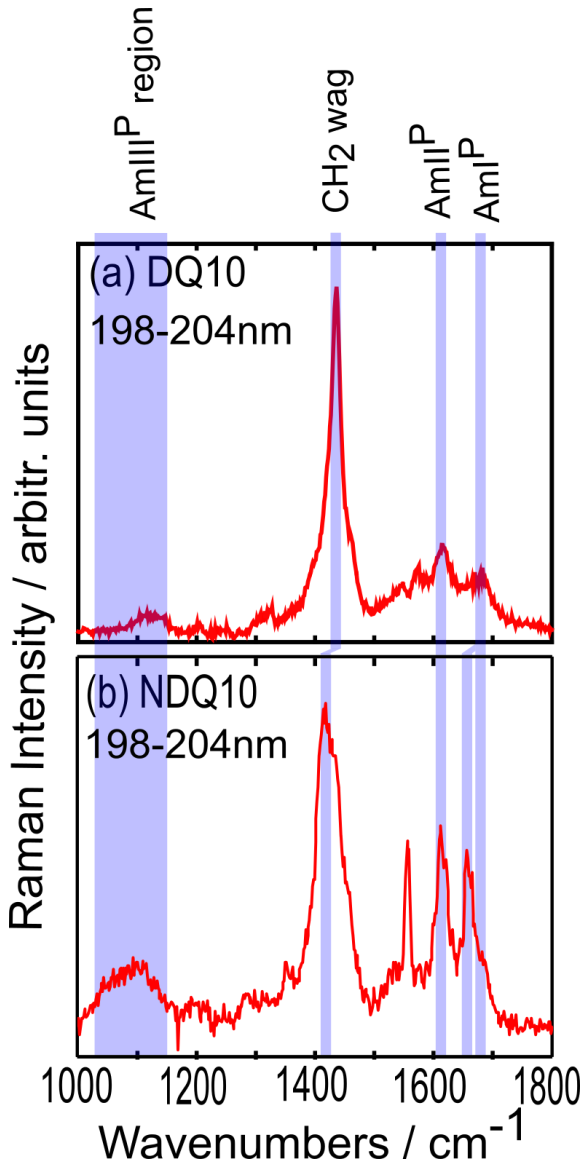

Figure 5.4: UVRR 198 nm – 204 nm difference spectra of Q10. The spectra are of (a) DQ10 reported by Xiong et al.[57] and (b) NDQ10 reported in this work. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

As shown in Table 5.1, the AmI^P and $AmII^P$ band frequencies of DQ10 are similar to those of monomeric glutamine in water. This indicates that the primary amide C=O and NH_2 groups of DQ10 are predominately hydrogen bonded to water. In contrast, the AmI^P band of NDQ10 is downshifted to ~1660 cm^{-1}, suggesting that a significant population of side chain C=O groups are involved in intrapeptide hydrogen bonding. Additionally, the shoulder at ~1680 cm^{-1} suggests that there is also a population of NDQ10 side chain C=O groups that hydrogen bond to water molecules.

Table 5.1: Glutamine, DQ10, and NDQ10 Primary Amide UVRR Band Frequencies in Water (cm^{-1}).

Vibration	Glutamine (cm^{-1})	DQ10 (cm^{-1})	NDQ10 (cm^{-1})
AmI^P	1679	1680	1680 (sh)
			1655
$AmII^P$	1620	1614	1614
CH_2 wag	1427	1437	1432 (sh)
			1414
$AmIII^P$	1110	1099	1106

sh: shoulder

In contrast, the $AmII^P$ band frequency of NDQ10 (~1614cm^{-1}) indicates that its side chain NH_2 groups predominately hydrogen bond to water. Thus, we conclude that NDQ10 side chain C=O moieties must be hydrogen bonded to the NH groups of the peptide backbone. This conclusion is supported by a recently published report by Walsh et al.[120] who showed that small, gas phase glutamine dipeptides with a β-strand-like conformation preferentially form hydrogen bonds between side chain C=O groups and backbone NH groups.

5.4.5 Glutamine Side Chain χ_3 Angle Distributions of DQ10 and NDQ10

We recently discovered a side chain UVRR primary amide vibration that shows a cosinusoidal frequency dependence on the O–C–C–C dihedral angles of glutamine and

asparagine side chains (designated as the χ_3 and χ_2 dihedral angles, respectively).[71] This vibration, which we call the AmIII^P, is reminiscent of the AmIII$^S{}_3$ vibration because it derives from an in-phase combination of side chain C-N stretching and NH_2 rocking motions (instead of peptide backbone C–N stretching and N–H bending of the AmIII$^S{}_3$). It should be noted however, that C_β–C_γ stretching also contributes to the AmIII^P vibration.

The structural sensitivity of the AmIII^P derives from hyperconjugation between the side chain C=O π^* and C_β–C_γ σ orbitals.[71] When hyperconjugation is strong (e.g. at $\chi_3 \sim \pm 90°$) electron density is transferred from the C_β–C_γ to the C=O bond. This elongates the C_β–C_γ bond and reduces the C_β–C_γ stretching force constant. As a result, the AmIII^P frequency downshifts. In the absence of hyperconjugation (e.g. at $\chi_3 \sim 0°$), the C_β–C_γ bond length shortens and the AmIII^P frequency upshifts.

We examined the AmIII^P bands of the Figure 5.4 UVRR difference spectra of NDQ10 and DQ10 to determine their χ_3 angle distributions (see Appendix A for details). As discussed previously by Punihaole et al.,[71] the AmIII^P frequency of DQ10 at ~1099 cm^{-1} correlates to a χ_3 distribution centered at $\sim -30°$ and/or $\sim 20°$ (Figure 5.3e). This result is similar to that of aqueous glutamine.[71] For NDQ10 (Figure 5.3g), the AmIII^P band peaks at ~1106 cm^{-1}, which correlates to χ_3 angle distributions that are centered at $\sim -12°$ and/or $\sim 3°$.

5.4.6 DQ10 and NDQ10 are Monomers in Aqueous Solution

We used diffusion-ordered NMR spectroscopy (DOSY) to measure the diffusion coefficients of DQ10 and NDQ10. Figure 5.5a and b show representative signal decays of NMR peaks corresponding to Gln side chain N-^{1}H resonance for DQ10 and NDQ10 respectively. We

fit these decays to eq. 5.1 and obtained diffusion coefficients of 1.91(7) × 10^{-10} m^2·s and 1.79(3) × 10^{-10} m^2·s for DQ10 and NDQ10, respectively. Assuming a sphere, we roughly estimate that both coefficients are consistent with diffusion of monomeric peptides with Stokes radii of ~13-14 Å. The fact that DQ10 and NDQ10 are monomers in solution is also supported by our UVRR hydrogen-deuterium exchange (HX) measurements (see Appendix A for details). As shown in Figure 5.5c, the HX exchange times of DQ10 and NDQ10 are essentially identical, ~5.5 min, consistent with other monomeric peptides.[121]

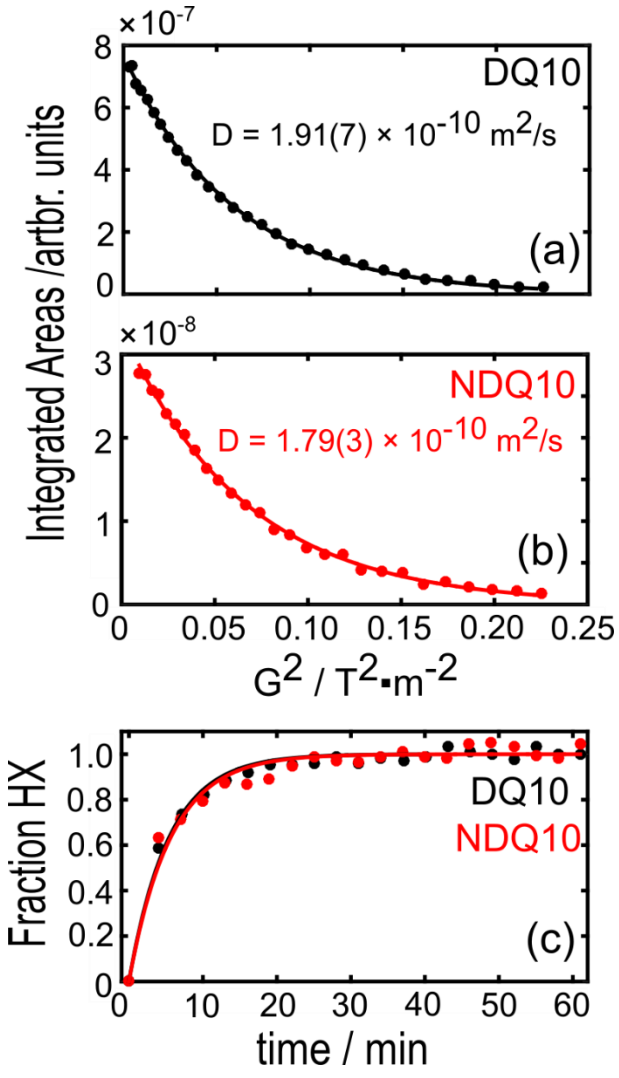

Figure 5.5: DOSY NMR and HX UVRR data for DQ10 and NDQ10. The data indicate that DQ10 and NDQ10 are monomers in solution. Representative DOSY NMR signal decays for (a) DQ10 and (b) NDQ10. The measured diffusion coefficients for DQ10 and NDQ10 are 1.91(7) × 10−10 m^2/s and 1.79(3) × 10−10 m^2/s, respectively. (c) HX kinetics of DQ10 and NDQ10 are similar, showing time constants of ~5.5 min, which are similar to those of monomeric peptides of similar molecular weights. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

5.4.7 Qualitative Evidence of a Large Activation Barrier between the DQ10 and the NDQ10 Solution Monomer Conformations

We found that the solution conformations of DQ10 and NDQ10 do not readily interconvert. Additionally, Xiong et al. [23] showed that the soft, possibly-crystalline solids, formed upon evaporating *DQ10*, retain their solution-state conformation. These observations suggest that the NDQ10 and DQ10 solution conformations lie in separate deep energy wells. The high activation barrier between NDQ10 and DQ10 effectively prevent conformational transitions and equilibration between their respective structures.

5.4.8 ~2-Fold Decreased H-bonding Solvent Transforms DQ10 from a PPII-like to an α-helix-like-conformation, while a >3-Fold Decreased H-bonding Solvent Transforms DQ10 to β-sheet Fibril Conformations.

We investigated the conformational transitions of DQ10 in less hydrogen bonding, lower dielectric constant acetonitrile/water solvent mixtures. The addition of acetonitrile, is expected to promote inter- and intramolecular peptide hydrogen bonding by reducing the solvent's ability to hydrogen bond.[122] Figure 5.6a shows the UVRR spectra of DQ10 in a series of water/acetonitrile mixtures. Compared to DQ10 in pure water, the relative intensity of the $AmIII^{S}_{3}$ band decreases as the mole fraction of water decreases to $X_{H2O} = 0.44$. At mole fractions below $X_{H2O} = 0.44$, the $AmIII^{S}_{3}$ band downshifts to ~1230 cm^{-1} and increases significantly in relative intensity. The dramatic spectral changes that occur at low mole fractions of water are accompanied by the formation of large, visually evident aggregates.

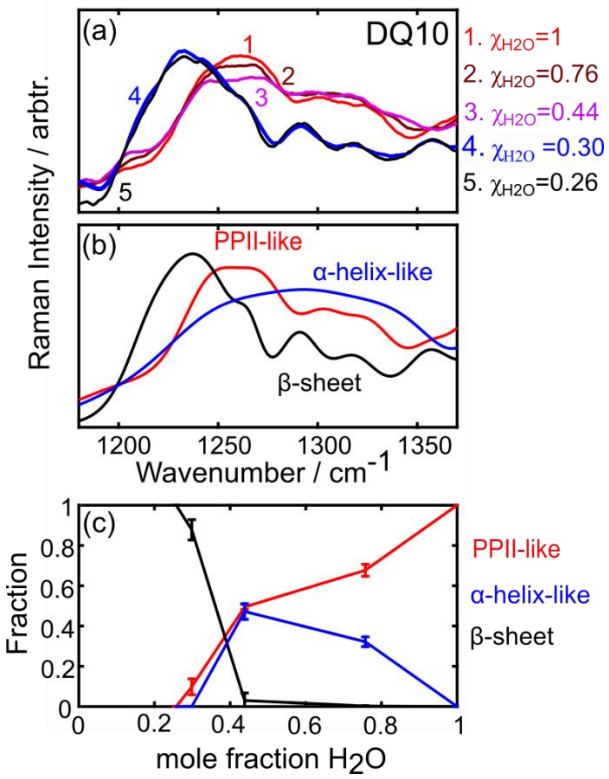

Figure 5.6: Dependence of the 204 nm UVRR Spectra of DQ10 on the H_2O/CD_3CN solvent composition. (a) Dependence of the Amide III band on solvent composition. The spectra were smoothed using a Savitzy-Golay filter with a second order polynomial over a 15 point spectral window. (b) Calculated Q10 β-sheet (black), PPII-like (red), and α-helix-like (blue) UVRR basis spectra (see Appendix A, Figure A.2 for details). (c) Relative fractions of PPII-like (red), α-helix-like (blue), and β-sheet (black) secondary structures as a function of H_2O mole fraction. The error bars were determined from three independent experiments. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

The spectra shown in Figure 5.6a were modeled as a linear combination of basis spectra to determine the relative fraction of different secondary structures. We found that the spectra could be satisfactorily modeled using a linear combination of PPII-like, α-helix-like, and β-sheet basis spectra (shown in Figure 5.6b). Our modeling procedure and methods used to calculate the basis spectra are discussed in detail in Appendix A (Figure A.1 and Figure A.2).

Figure 5.6c shows the relative fractions of the different secondary structure conformations of DQ10 as a function of the acetonitrile/water solvent composition, while Figure 5.7 shows the consequent Ψ angle distributions. The UVRR data indicate that the DQ10 peptide bonds adopt both PPII-like and α-helix-like conformations (Figure 5.1c) at intermediate water/acetonitrile mole fractions (X_{H2O} = 0.76 or X_{H2O} =0.42), while at lower mole fractions of water (χ_{H2O} = 0.24) DQ10 forms β-sheet-rich aggregates (Figure 5.1d). These data agree with our measured CD spectra shown in Figure A.3 of Appendix A.

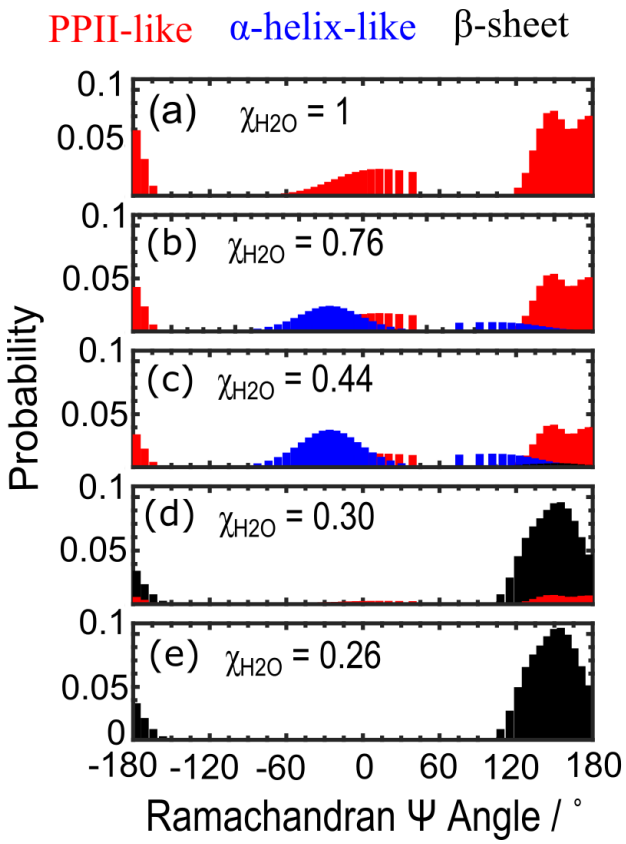

Figure 5.7: Ramachandran Ψ Angle Distributions of DQ10 in different H_2O/CD_3CN Mixtures. The distributions shown in red correspond to PPII-like structures, in blue to α-helix-like structures, and in black to β-sheet structures. The Ψ angle distributions were determined from the Figure 5.6b basis spectra (see Appendix A for details) and scaled according to the relative fractions of PPII-like, α-helix-like, and β-sheet structures reported in Figure 5.6c. Reprinted with permission from Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

The distributions shown in Figure 5.7b and c that derive from peptide bonds in "α-helix-like" conformations (shown in blue) possess two broad populations of Ψ angles centered at −25° and 100°. The distributions indicate that the peptide bonds do not derive from canonical α-helix conformations. Instead, they suggest that a large fraction of these structures likely derive from 3_{10}-like helices (indicated by the distribution peak at −25°), interspersed by turn structures (indicated by the peak located at ∼100°). However, the breadth of these distributions indicates that the DQ10 peptide bonds also populate Ψ angles indicative of more canonical α-helix conformations, as well as small populations of π-helices.

In addition to dominant α-helix-like and PPII-like conformations, a small population of peptide bonds adopt β-sheet-like structures in acetonitrile/water mixtures where X_{H2O} = 0.44. However, at lower water mole fractions (χ_{H2O} = 0.30), the PPII-like and α-helix-like structures convert almost entirely into β-sheet structures (Figure 5.6c and Figure 5.7d, e). The formation of these β-sheet structures coincides with the appearance of visually evident aggregates.

TEM images (Figure 5.8) indicate that these β-sheet aggregates consist of oligomers and mature amyloid-like fibrils. At χ_{H2O} = 0.44 (Figure 5.8a), we observe both fibrillar and oligomeric aggregates; however, at χ_{H2O} = 0.26 (Figure 5.8b), we observe essentially only fibrils. Replacement of water with increasing amounts of acetonitrile destabilizes PPII-like structures because the solvent is less competitive for hydrogen bonding to the backbone peptide bonds. As a result, Q10 adopts α-helix-like and turn conformations, which can satisfy hydrogen bonding through intramolecular interactions. However, at high concentrations of acetonitrile, hydrogen bonding must involve inter-peptide chain hydrogen bonding, thus forming aggregates and β-sheet-rich fibrils.

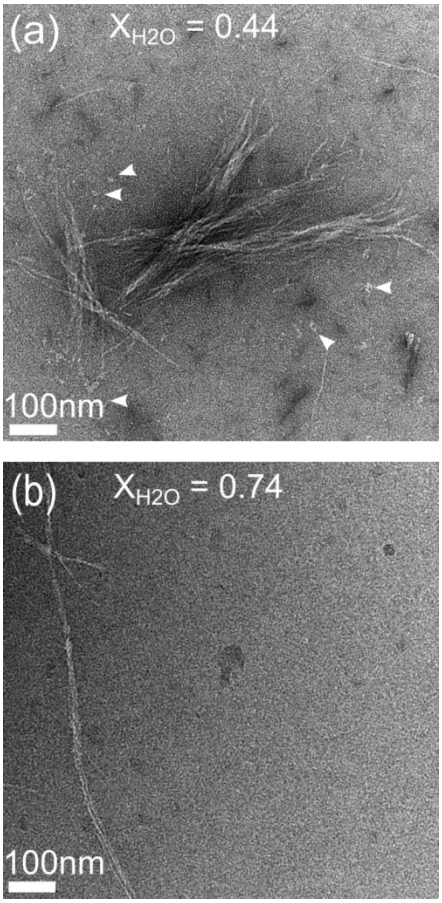

Figure 5.8: Representative TEM images of DQ10 aggregates. Aggregates were prepared in (a) X_{H2O} =0.44 solution and (b) X_{H2O} = 0.26 solution. The white arrows point to non-fibrillar aggregates. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

The morphologies and UVRR spectra of the DQ10 fibrils prepared at low mole fractions of water are similar to fibrils prepared by Punihaole et al.[46] in aqueous NDQ10 and DQ10 monomer solutions at 60°C and pH 7 (

Figure 5.1e). Punihaole et al. found that the fibrils prepared from both DQ10 and NDQ10 peptide solutions in this manner were composed of extended β-strands that predominately assembled into antiparallel β-sheets, although a minor population of parallel β-sheets were also observed. Here, we also observe that ~95% of the peptide bonds adopt antiparallel β-sheet conformations in fibrils prepared in acetonitrile/water mixtures.

5.4.9 Fibrils Aggregated in Low H-bonding Solvent Re-dissolve in Water into the NDQ10 Monomer Solution Conformation.

In contrast to fibrils prepared from DQ10 monomers by Punihaole et al. [46], those prepared from the acetonitrile solutions re-dissolve in water (

Figure 5.1a and d). These re-dissolved aggregates show a UVRR spectrum identical to the solution-state, monomeric NDQ10 conformations (see Figure A.4). This result significantly differs from fibrils formed via incubation in water, which irreversibly form fibrils.

We investigated the side chain and backbone amide hydrogen bonding in DQ10 fibrils grown in acetonitrile by examining their 197–204 nm and 204-(197-204) nm UVRR difference spectra respectively (Figure A.8 and Figure A.9). We observe an AmI^S frequency at ~1670 cm^{-1} and an AmI^P frequency at ~1663 cm^{-1}. In contrast, DQ10 fibrils grown in water have AmI^P and AmI^S frequencies of 1664 cm^{-1} and 1661 cm^{-1} respectively.[46] The $AmIII^S{}_3$ band of fibrils prepared in acetonitrile is essentially identical to that grown in water.[46] This suggests that the

peptide backbone structures of these fibrils are essentially the same. Thus, we conclude that the AmI^P and AmI^S frequency differences must be attributed to changes in hydrogen bonding of the side chain and backbone amide groups respectively. Based on this conclusion, our data suggest that, compared to aqueous solution, fibrils prepared from DQ10 in acetonitrile have similar side chain C=O and NH_2 hydrogen bonding strengths, but much weaker peptide backbone hydrogen bonding strengths. The weakened backbone hydrogen bonding interactions may destabilize the fibril structure causing fibrils grown in acetonitrile to dissolve when exposed to more water.

Interestingly, DQ10 fibrils grown in ~90% acetonitrile and re-dissolved in 50% acetonitrile/50% water show CD spectral signatures characteristic of NDQ10 collapsed β-strands, with no indication of α-helix-like conformations (Figure A.5). At intermediate acetonitrile concentrations, DQ10 PPII-like structures are in equilibrium with α-helix-like conformations. These α-helix-like structures convert back into PPII-like conformations upon further dilution with water (Figure A.5).

At high acetonitrile mole fractions (χ_{H2O}=0.26), NDQ10 shows a Raman spectrum characteristic of β-sheet fibrils; however, unlike DQ10, these solutions are not turbid. Upon centrifugation (~20,000 × g), solutions of NDQ10 with high acetonitrile mole fractions do not form pellets. Thus, it appears that NDQ10 solutions with high acetonitrile mole fractions do not form high molecular weight aggregates. Copious water addition to NDQ10 in 90% acetonitrile, results in a conformational transition to its original structure (Figure A.10).

5.4.10 Metadynamics Simulations of Q10

We used metadynamics MD simulations to obtain insights into the conformational equilibrium and the high activation barrier between the DQ10 and NDQ10 monomer forms. Metadynamics simulations are ideal for characterizing the conformational landscapes of complex biophysical systems.[123,124] In metadynamics, the evolution of the simulated system is biased by adding a history-dependent potential energy that discourages revisiting events already sampled and accelerates the sampling of rare events. As a consequence of the algorithm, metadynamics can also be utilized to estimate the conformational free energy landscape of a peptide or protein as a function of a suitably chosen set of reaction coordinates called collective variables.

Figure 5.9 shows the conformational Gibbs free energy landscape of Q10 as a function of three collective variables, defined in terms of the root-mean-square-deviations (RMSD) from idealized β-hairpin, α-helix, and PPII structures. We chose these three reference structures because they are common secondary structure motifs that enable us to clearly distinguish between different possible polyQ conformations.[125–127] We chose a β-hairpin structure because it was previously implicated as an aggregation nucleus.[128]

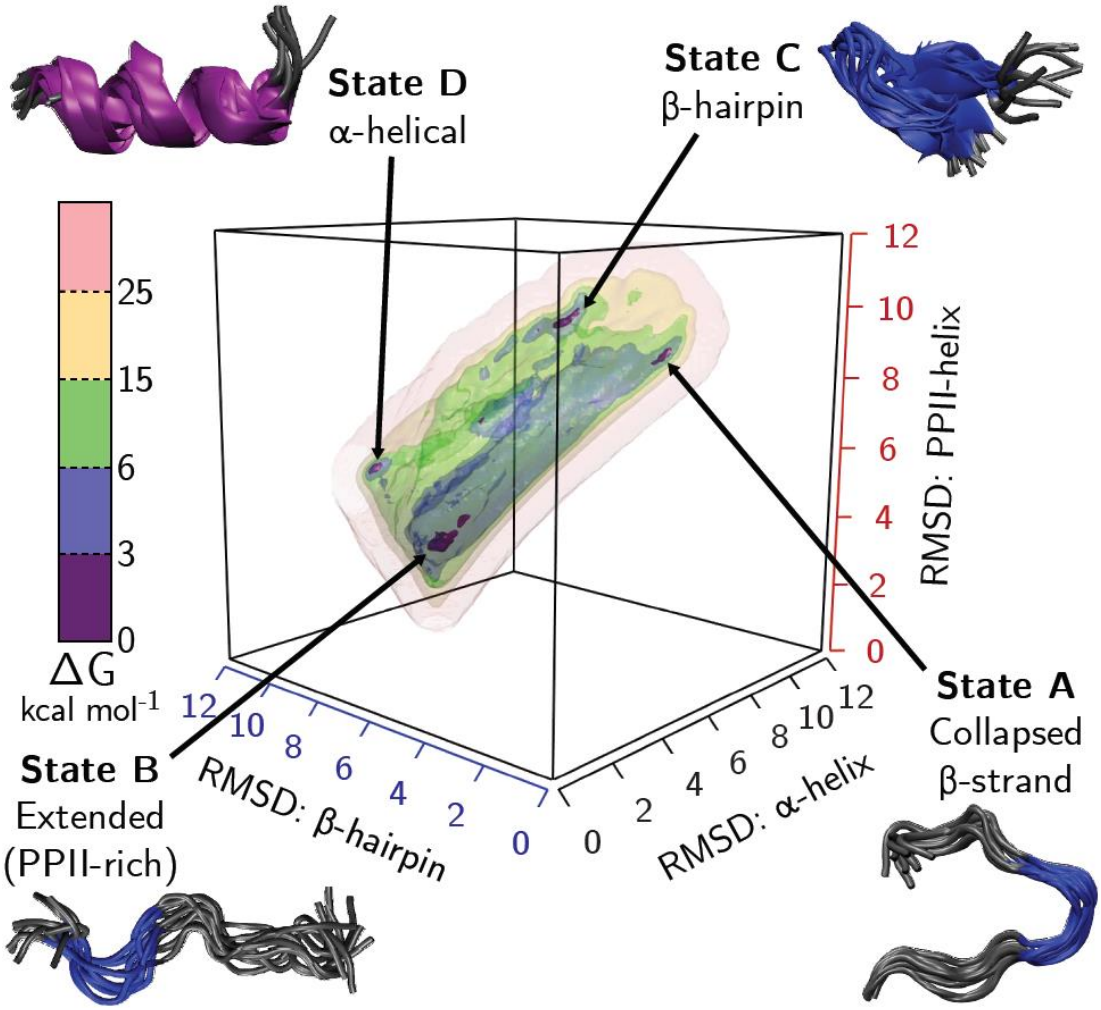

Figure 5.9: Conformational energy landscape of Q10 obtained from metadynamics. The four lowest energy states are represented by the violet isosurfaces with representative structural ensembles shown for these low energy states. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967.

We find that the conformational energy landscape of Q10 is frustrated because it contains numerous shallow, local minima. However, four deep energy minima (purple regions in Figure 5.9) are observed. These energy minima contain different structural states (A–D) that are defined both by their Ramachandran angles (Figure 5.10a–d), as well as their hydrogen bonding propensities (Figure 5.11).

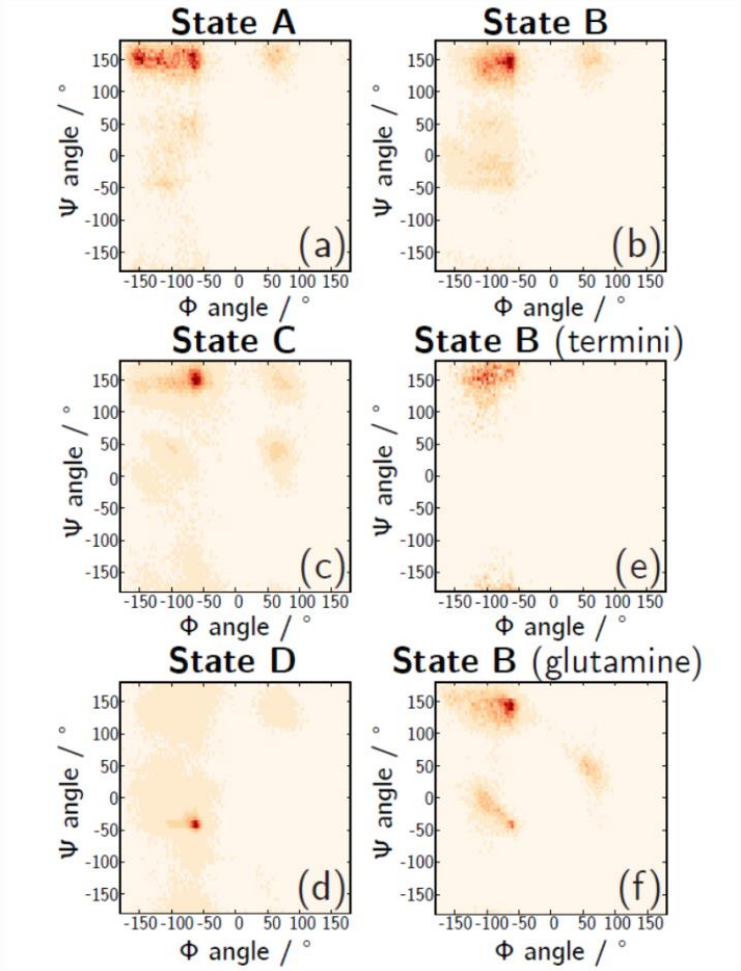

Figure 5.10: Ramachandran (Φ, Ψ) angle distributions for metadynamics structure of Q10. The Ramachandran plots are of the following: (a) collapsed β-strand structures of state A, (b) PPII-like structures of state B, (c) β-hairpin-like structures of state C, and (d) α-helical structures of state D. Panel (e) shows the (Φ, Ψ) angles of the terminal aspartic acid and lysine residues of state B while panel (f) shows (Φ, Ψ) angles of the glutamine residues for state B. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society

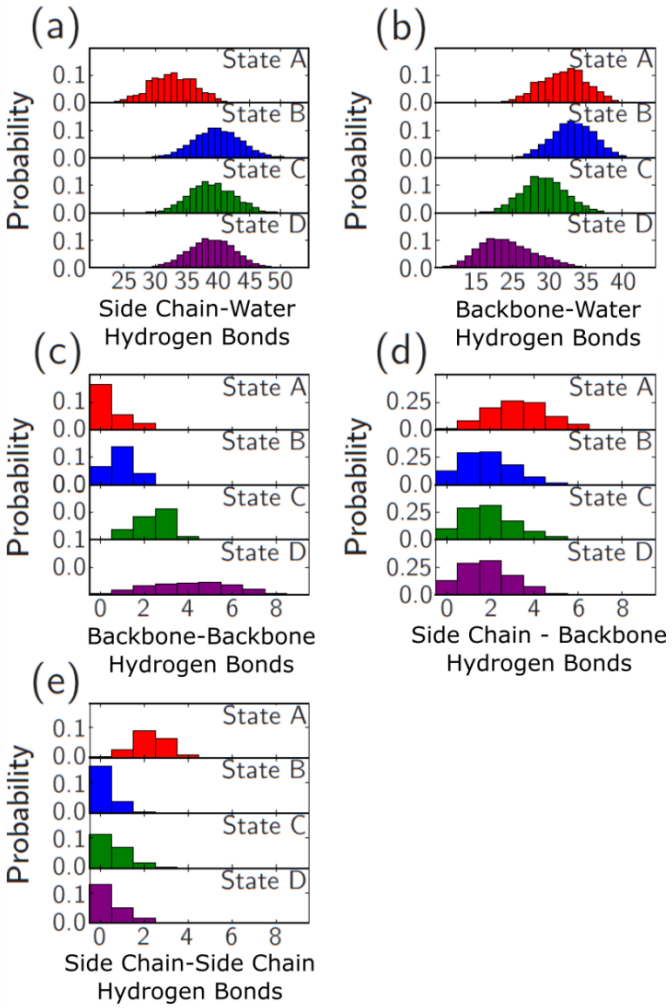

Figure 5.11: Number of inter-amide and amide-water hydrogen bonds formed for metadynamics structures. (a) Total number of glutamine side chain-water hydrogen bonds for each structure; (b) total number of peptide backbone-water hydrogen bonds formed for each structure; (c) total number of peptide backbone-backbone hydrogen bonds formed for each structure; (d) total number of side chain-backbone hydrogen bonds for each structure; (e) total number of side chain-side chain hydrogen bonds formed for each structure. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society

The structures found in state A possess peptide bonds with average (Φ, Ψ) angles of (−90°, 140°), roughly consistent with *β*-strand-like conformations (Figure 5.10a). These conformations are best described as "collapsed" *β*-strands, rather than *β*-hairpin, because they lack the intramolecular peptide backbone hydrogen bonding that defines canonical *β*-hairpins. As shown in Figure 5.11d, these *β*-strand structures form, on average, more side chain-backbone hydrogen bonds than the other states. These side chain-backbone hydrogen bonds, together with the attractive electrostatic interactions between the terminal aspartic acid and lysine residues, cause the *β*-strands to adopt more compact (i.e. collapsed) conformations**.**

The (Φ, Ψ) angle distributions of state B are centered at about (−80°, −150°), which is characteristic of PPII secondary structures (Figure 5.10b). However, a smaller population of Ramachandran angles, located at approximately (−50°, −57°), indicate that state B also contains turn-like conformations. Taken together, these two populations of Ramachandran angles indicate that state B consists of irregular structures that are characterized by short PPII helices interspersed by turn conformations or bends. In addition, our hydrogen bonding analysis (Figure 5.11) indicates that, compared to collapsed β-strand structures (state A), both the peptide backbone and side chains of PPII-like structures (state B) are preferentially hydrogen bonded to water.

The structures found in state C possess peptide bonds whose (Φ, Ψ) angles are centered at (−135°, 135°), indicative of *β*-strand structures (Figure 5.10c). A smaller population of peptide bonds adopts Type I' turn Ramachandran angles (−40°, −85°). In contrast to the state A collapsed *β*-strand structures, the state C *β*-hairpin structures are calculated to form more hydrogen bonds between backbone peptide bonds (Figure 5.11c). These backbone-backbone hydrogen bonds are between *β*-strands and are characteristic of canonical *β*-hairpins defined by Milner et al.[129] In

addition, the side chains of these β-hairpin structures are more solvated than those of the collapsed β-strand structures found in state A (Figure 5.11a).

The structural ensemble of state D is comprised mostly of α-helical conformations. As shown in Figure 5.10e, the (Φ, Ψ) angles are distributed narrowly around (−50°, −65°). These structures show substantial amount of backbone-backbone hydrogen bonding (Figure 5.11c), as expected for α-helices, while the side chains are well solvated.

The collapsed β-strand (State A) occurs at the global minimum in the conformational energy landscape, while the extended PPII conformation (state B), the β-hairpin (state C), and the α-helix (State D) occur at 0.5 kcal · mol^{-1}, 0.66 kcal · mol^{-1}, and 1.0 kcal · mol^{-1} higher in energy, respectively. A 3-6 kcal · mol^{-1} activation energy barrier separates States A and B. In contrast, states C and D are *both* isolated from each other and isolated from states A and B by a 6-15 kcal · mol^{-1} energy barrier.

5.4.11 Consistency of MD Simulation and UVRR Results

Our interpretation of the Ramachandran Ψ and Φ angle distributions for the simulated states A-D (Figure 5.10a–d) indicate that the state A (collapsed β-strand) and state B (PPII-like) structures resemble the UVRR measured solution structures found for the NDQ10 and DQ10 monomers, respectively. To validate the simulation results, we compared the UVRR experimentally measured Ψ and χ_3 angles of DQ10 and NDQ10 to the corresponding distributions obtained from the MD structures populating the states A and B energy wells (Figure 5.3). Overall, agreement between the experimentally measured Ψ angle distributions and those

obtained from the simulations is excellent. In contrast, the side chain χ_3 angle distributions obtained from the MD simulation show poor agreement with the UVRR data.

Figure 5.3a–d compare the Ψ angle distributions of DQ10 and NDQ10 with those of states A and B from the simulations. The corresponding experimentally measured and computationally derived Ψ angle distributions are in overall good agreement. The only major discrepancy is that the calculated distribution of state B (Figure 5.3b) lacks a peak at ∼175° that corresponds to 2.5_1-helix conformations. This apparent discrepancy is discussed in detail below.

Metadynamics show that the χ_3 distributions for states A and B are bimodal (Figure 5.3f and h). Both distributions show peaks centered at ∼−70° and ∼110°. These χ_3 angle distributions differ significantly from those measured experimentally by UVRR. An analysis of a side chain rotamer database by Dunbrack and Shapovalov[23,71] shows that glutamine residues typically possess χ_3 angles that are distributed between ∼−90 to 90°, which is inconsistent with the distributions shown in Figure 5.3f and h.

We conclude that the large differences between our UVRR measurements and our MD simulated χ_3 angle distributions derive from a lack of parameterization for non-rotameric side chain dihedral angles in the CHARMM36 force field. It is known, for example, that the χ_1 and χ_2 dihedral angles of amino acid side chains have been optimized in CHARMM36 using experimental data.[130] However, the non-rotameric dihedral angle, such as the χ_3 angle of the glutamine side chain, have not been similarly treated in CHARMM36.

We compared the backbone and side chain amide hydrogen bonding environments observed by UVRR measurements (NDQ10 and DQ10) to those calculated from the metadynamics calculated structures of states A and B (Figure 5.11). Our computational results are in excellent agreement with our experimental findings. The state B structures of the

metadynamics contains side chain and backbone peptide bonds predominantly hydrogen bonded to water, while the state A structures show a significant inter-amide hydrogen bonding.

The metadynamics data validate the observation of a large activation barrier between the NDQ10 PPII (state B) and the DQ10 β -strand (state A) conformations. We calculate that the barrier between states A (β-strand) and B (PPII) is 3-6 k_BT. In addition, the simulated MD energy barrier region between states A and B also contains ≥ 8 local energy minima, with energy well depths of ~ 1.0 kcal · mol^{-1} (Figure A.6). These local energy minima define metastable states along the energy barrier region. Even if Q10 can overcome the large energy barrier, the kinetics of conformation equilibration will be slowed by these multiple wells to prevent PPII-like to *β*-strand structural transitions.

5.4.12 Origin of Q10 2.5_1-helix Conformation Stabilization

Our DQ10 UVRR studies surprisingly detect a significant population of the 2.5_1-helix conformation previously proposed by Krimm and coworkers[131–133], and observed for PLL and PGA peptides by Asher and coworkers.[119] The 2.5_1-helix conformation consists of an extended, left-handed, *β*-strand-like structure with ~ 2.5 residues per helical turn and (Φ,Ψ) angles of roughly (−130°, 170°).

Asher and Krimm's work indicate that the 2.5_1-helix conformations in PLL and PGA derive mainly from the electrostatic repulsions between their charged amino acid side chains forcing them to splay out.[119] Based on this understanding, it is surprising that a predominately polyQ-rich peptide such as Q10 (which only has four charged residues) would show a significant

population of peptide bonds with 2.5_1-helix-like Ψ angles (Figure 5.3a). Another disquieting fact is that our MD simulations of the state B structures (Figure 5.3b) appear unimodal, consisting of Ψ angles indicative of the PPII conformation. This lack of predicted 2.5_1-helix conformations is troubling because it could signal misinterpretations of the UVRR data.

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/2.5_1-helix dihedral Ramachandran 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 5.2). Neglecting the influence of solvating waters, our simulations indicate that the potential energy of the 2.5_1-helix is $\sim$50 kcal $\cdot$ mol^{-1} lower than that of the PPII structure, mainly due to intrapeptide electrostatics.

Table 5.2: Potential Energies (kcal·mole-1) for Q10 Peptides in PPII-like and 2.51-helix Conformations.

	PPII		PPII/2.5_1-helix		2.5_1-helix	
Energy Terms	all residues	glutamine residues	all residues	glutamine residues	all residues	glutamine residues
bonds	44(5)	33(5)	44(5)	32(5)	43(5)	32(5)
angles	134(10)	96(8)	134(10)	95(8)	134(10)	96(8)
torsions	146(5)	107(4)	160(5)	121(4)	174(6)	131(4)
van der Waals	−14(4)	−11(3)	−16(4)	−14(4)	−15(4)	−14(4)
electrostatics	−436(29)	−318(10)	−461(35)	−331(8)	−487(37)	−330(9)

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 structures is only ∼10 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 (Φ, Ψ) angle distributions of the glutamine residues and the charged flanking residues for state B structures. As shown in Figure 5.10e, the terminal aspartic acid and lysine residues preferentially adopt (Φ, Ψ) angles centered roughly at (−115°, −160°), closer to the (Φ, Ψ) angles of the canonical 2.5_1-helix conformation. In contrast, the glutamine residues (Figure 5.10f) adopt (Φ, Ψ) angles of (−80°, −150°) and (−50°, −57°), 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.

Thus, the UVRR DQ10 results appear consistent with the MD simulations. Assuming equal Raman cross sections, we roughly estimate from the integrated intensities of the $AmIII^{S}_{3}$ bands (Figure 5.2a) that 43% of peptide bonds adopt PPII-like Ψ angles, 22% of peptide bonds adopt 2.5_1-helix Ψ angles, and 35% of the peptide bonds adopt turn structure Ψ angles. Based on these percentages, we calculate that a *single DQ10 peptide* has 2–3 peptide bonds involved in a 2.5_1-helix conformation. These 2.5_1-helix conformation peptide bonds most likely occur on the terminal charged residues. However, the MD simulations indicate that the state B 2.5_1-helix-like (Φ, Ψ) angles are centered around −160°, very close to that of canonical PPII structures centered at −150°. This significantly differs from the Ψ = −175° value measured for DQ10 by UVRR, and causes the calculated Ψ angle distribution of state B (Figure 5.3b) to appear unimodal.

We attribute the discrepancy in the Ψ 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.[134] 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 parameterize CHARMM36. This conclusion is supported by Liqi Feng who showed that metadynamics of poly-l-lysine, using the CHARRM36 force field, fail to show a Ψ angle distribution at ~170°, indicative of 2.5_1-helix conformations.[135]

5.4.13 Comparisons to Other Studies

The solution-state structures of polyQ-rich peptides and proteins have been studied in detail.[37,43,45,99,136,137] Many of the important computational studies were conducted by Pappu's group.[37,45,99] For example, Wang et al.,[37] characterized the conformational ensemble of N-acetyl-Q_5-N'-methylamide (Q5) and N-acetyl-Q_{15}N'-methylamide (Q15), that are somewhat similar to our Q10 peptide.

From their MD simulations, Wang et al.[37] 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, β-strands, and α-helices occurring. They also show that both Q5 and Q15 peptide bonds have a strong propensity to adopt PPII-like and α-helix-like (Φ, Ψ) angles.

Wang et al's.[37] simulations also suggest that structurally disordered, monomeric polyQ peptides form a significant number of inter-amide hydrogen bonds, with side chain-backbone hydrogen bonding being the most prevalent. This is *ostensibly* supported by the NMR measurements by Darnell et al.[138] that find nuclear Overhauser effect (NOE) cross peaks between the C_γ protons of glutamine side chains and C_α protons of the backbone in R_3GQ_3GY and $R_3GQ_3P_{11}GY$ peptides that adopt predominately PPII structures. Darnell et al. interpret these signals to mean that in PPII-rich polyQ peptides, the glutamine side chains adopt "folded-over" conformations (see Figure 5.12) that enable hydrogen bonding to the backbone amides. To our

knowledge, this is the only experimental evidence of significant side chain-backbone hydrogen bonding in PPII-rich polyQ peptides.

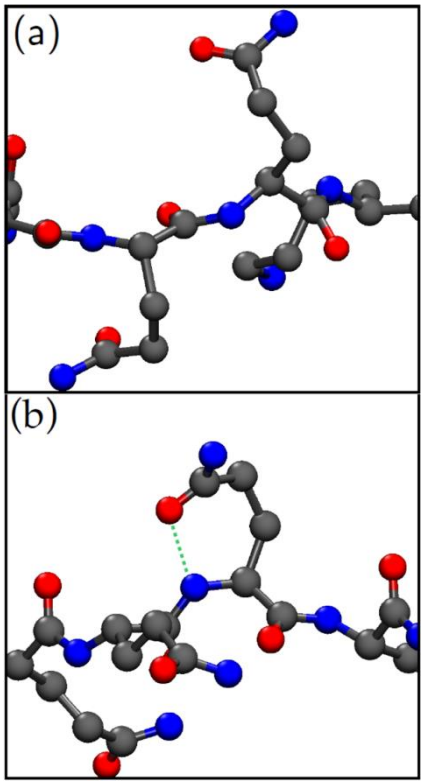

Figure 5.12: Proposed polyQ side chain conformations. (a) shows Q10 in a PPII conformation with extended, solvated side chains and (b) shows side chains hydrogen bonded to the peptide backbone as proposed by Darnell et al.[138] The structure shown in panel (a) is of the PPII state (State B) found in our metadynamics simulations and validated by our UVRR measurements of DQ10. The structure shown in panel (b) was created using a Q10 peptide with Ramachandran angles constrained to PPII values in which the side chain structure was modified to adopt side chain-backbone hydrogen bonds between the C=O and NH moieties of the side chains and backbone, respectively. We clearly do not experimentally observe significant side chain-backbone hydrogen bonding in PPII-like DQ10 peptide monomers. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society

The simulations of Wang et al.[37] lead them to hypothesize that the structural disorder of polyQ peptides stems from the many different possible combinations of intramolecular and intermolecular-amide hydrogen bonds that can form between side chain and backbone amides. This explains why a β-sheet-rich fibril nucleus[58,139] is expected to be energetically unfavorable. According to their model, the multiple possible combinations of intra-peptide hydrogen bonds promotes 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 et al.[37] 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 intramoleuclar side chain-side chain, backbone-backbone, and side chain-backbone hydrogen bonding interactions.

However, there are many aspects of our results that are inconsistent with those of Wang et al.[37] 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 5.1. Interestingly, we also find that the side chains and backbone amides of the structurally "disordered" PPII-rich state of Q10 are predominately hydrogen bonded to water, which disagrees with the Wang et al.[37] results.

These disagreements raise two important questions. First, why does Wang et al.[37] not observe *any* well-defined structural states for small polyQ peptides, whereas we do? And second, why do the simulations of Wang et al.[37] indicate that there are significant numbers of inter-amide hydrogen 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 enable us to robustly sample a greater ensemble of structures than does traditional MD simulations. Thus, we characterized the entire conformational landscape and discovered global minimum energy conformations.

We are also aware that the conformational behavior of model polyQ peptides in solution depends both on the number of glutamine repeats and the choice of non-glutamine flanking groups.[140] For example, the use of N-Acetyl and methylamide flanking groups in Pappu and coworkers' simulations are better models for polyQ tracts in proteins. In addition, fluorescence correlation spectroscopy[43] data by the Pappu group and dynamic light scattering data by the Murphy group[44] indicate that structurally disordered long polyQ peptides adopt relatively compact structures in aqueous solution. These compact structures presumably derive from the fact that inter-amide hydrogen bonding interactions become more prevalent and important in polyQ peptides as the glutamine repeat length increases. For example, Walters and Murphy[44] have shown that phosphate buffer is a good solvent for Q8 and Q12, a theta solvent for Q16, and a poor solvent for larger peptides such as Q20. Thus, the model proposed by Pappu and coworkers in the Wang et al.[37] study may be valid in the limit of polyQ peptides with ≥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 Wang et al.[37] As discussed by Wang et al.,[37] the hydrogen bonding interactions observed in a simulation will vary depending on the force field utilized. Our

use of the TIP3 water model and the CHARMM36 modern force-field results in different hydrogen bonding interactions compared to that of the MD simulations of Wang et al.[37] The hydrogen bonding interactions predicted by our metadynamics are in excellent agreement with our UVRR experimental data, which shows that our choice of force-field appropriately models the hydrogen 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 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 β-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 over-stated, *at least in the context of small peptide systems, such as Q5, Q10, and Q15*.

5.5 Conclusions

We investigated the structural ensemble of the Q10 polyQ peptide in solution. From our UVRR, CD and MD data, we constructed a molecular-level model for the structure and part of the energy landscape of DQ10 and NDQ10 monomers in aqueous solution. We show that, in DQ10, the charged, terminal residues adopt 2.5_1-helix conformations, while the central glutamine residues adopt short PPII segments that are interspersed with turn structures (Figure 5.1b). In contrast, we find that NDQ10 exists in a collapsed β-strand conformation (Figure 5.1a).

There is a large activation barrier between these conformations in solution that prevents their interconversion.

Our study achieves new insights into the hydrogen bonding environments and glutamine side chain conformations of Q10. The DQ10 side chains are predominately hydrogen bonded to water, whereas the NDQ10 side chains form relatively more inter-amide hydrogen bonds. We were also able to determine that DQ10 possesses χ_3 angle distributions centered at $\sim -30°$ and/or $\sim 20°$. In contrast, NDQ10 has a χ_3 angle distribution centered at $\sim -12°$ and/or $\sim 3°$.

We also monitored the PPII to β-sheet structural transition of DQ10 in acetonitrile/water mixtures. Acetonitrile is a secondary structure inducing solvent because it promotes intra- and inter-peptide hydrogen bonding. At intermediate acetonitrile concentrations, we observe a significant amount of α-helix-like secondary structures in DQ10 (Figure 5.1c). At high acetonitrile concentrations, we observe a transition to β-sheet conformations and the formation of visible aggregate light scattering (Figure 5.1d). The higher acetonitrile concentrations promote intra- and inter-peptide hydrogen bonding by first forming α-helix-like structures, and at higher acetonitrile concentrations forming β-sheet aggregates that include fibrils. The DQ10 fibrils and β-sheet aggregates formed in acetonitrile/water solutions re-dissolve upon addition of water. We find that the re-dissolved fibrils have a UVRR spectrum identical to monomeric NDQ10 indicating a collapsed β-strand structure. To our knowledge, this is the first report of re-dissolvable polyQ fibrils.

In contrast to DQ10, NDQ10 at high acetonitrile concentrations does not become turbid and does not form a pellet upon centrifugation. Thus, we conclude that NDQ10 does not form high molecular weight aggregates at high acetonitrile concentrations. However, the UVRR spectra of NDQ10 at high acetonitrile mole fractions are similar to that of fibrils and show a Ψ

angle distribution characteristic of β-sheet. This NDQ10 collapsed β-strand to β-sheet transition upon the addition of acetonitrile is reversible.

We also used metadynamics to investigate the energy barriers between the collapsed β-strand and the DQ10 PPII/2.5_1 helix conformation. We calculate a 3-6 k_BT activation barrier with additional smaller minima along the reaction coordinate between these conformations (Figure 5.1a and b). This is consistent with our experimental observation that NDQ10 and DQ10 conformations do not readily interconvert.

Another important result of this study is that we rigorously bench-marked the CHARMM36 force field against experimental data and determined both the strengths and limitations of this force field's accuracy. We will continue to investigate and improve the utility of CHARMM36 in describing 2.5_1-helix structures, as well as the non-rotameric χ_3 dihedral angles calculated for glutamine side chains.

Until now investigations of the aggregation mechanisms of polyQ peptides and proteins have been hindered by the inability to study intermediate polyQ species along the aggregation and fibril formation reaction coordinate. The fact that we can poise our Q10 peptide in different conformations and reversibly transition these conformations along reaction coordinates that form β-sheet aggregates, such as fibrils, makes us hopeful that this simple Q10 system will enable elucidation of fibril aggregation mechanisms. We will be examining polyQ structural changes in physiologically relevant polyQ peptides in future studies.

5.6 Associated Content

Supporting Information is in Appendix A. Supporting Information includes further information on UVRR spectral processing and fitting, determination of Ψ and χ3 angle distributions, hydrogen-deuterium exchange experiments, and CD measurements.

5.7 Acknowledgements

Funding for this work was provided by the University of Pittsburgh (DP, RSJ, SAA), Defense Threat Reduction Agency HDTRA1-09-14-FRCWMD (RSJ, SAA), and partially supported by NIH R01 DA027806 (JDM and RJW). The MD simulation computer time was supported by XSEDE MCB060069, and computer equipment was purchased from NSF funds (CHE-1126465 and P116Z080180).

6.0 Interaction Enthalpy of Side Chain and Backbone Amides in Polyglutamine Solution Monomers and Fibrils

Adapted with permission from: David Punihaole, Ryan S. Jakubek, Riley J. Workman, and Sanford A. Asher. "Interaction Enthalpy of Side Chain and Backbone Amides in Polyglutamine Solution Monomers and Fibrils." The journal of physical chemistry letters, 2018, 9, 8 1944-1950.

Author Contributions: DP and RSJ contributed equally to this work. DP and RSJ analyzed the Raman data and developed the theory. RSJ developed error estimations for experimental results with assistance from the University Of Pittsburgh, Department Of Statistics. RJW performed computational work and analyzed it with the assistance of DP and RSJ. The manuscript was prepared by DP, RSJ, and SAA.

We determined an empirical correlation that relates the Amide I vibrational band frequencies of the glutamine (Q) side chain to the strength of hydrogen bonding, van der Waals, and Lewis acid-base interactions of its primary amide carbonyl. We use this correlation to determine the Q side chain carbonyl interaction enthalpy (ΔH_{int}) in monomeric and amyloid-like fibril conformations of $D_2Q_{10}K_2$ (Q10). We independently verified these ΔH_{int} values through molecular dynamics simulations that showed excellent agreement with experiments. We find that side chain-side chain and side chain-peptide backbone interactions in fibrils and monomers are more enthalpically favorable than are Q side chain-water interactions. Q10 fibrils also show a more favorable ΔH_{int} for side chain-side chain interactions compared to backbone-backbone interactions. This work experimentally demonstrates that inter-amide side chain interactions are important in the formation and stabilization of polyQ fibrils.

6.1 Introduction

Ten neurodegenerative diseases, including Huntington's, are linked to mutational expansions of polyglutamine (polyQ) repeats in proteins.[1] The increase in polyQ repeats greatly enhances misfolding and aggregation of affected proteins. Although the exact mechanism of neurotoxicity is still heavily debated, the pathological hallmark of all these diseases is the formation of neuronal aggregates composed of β-sheet-rich amyloid-like fibrils.[33,141,142] Given their potential role in neurotoxicity, there is great interest in understanding polyQ fibril formation, as well as developing therapeutic strategies to inhibit aggregation.

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 Q 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.[34–36,41,44,143] These findings are also supported by computational studies,[37–39,99] which suggest that polyQ peptides are largely disordered due to the multiplicity of different hydrogen bonding interactions that form between side chain and backbone amides. Other computational studies suggest that inter-amide hydrogen bonds between side chains contribute most significantly to the structural stability of polyQ amyloid-like fibrils.[144]

These and other studies[1,145] underscore the crucial role that Q side chain hydrogen bonding interactions play in dictating the solution-state conformational behavior and the strong aggregation propensities of polyQ peptides. Surprisingly, however, no experimental studies have quantified the relative energetic favorability of side chain versus backbone amide hydrogen bonding interactions in polyQ peptides. Thus, developing new experimental tools that can

quantify the relative energies of different side chain and backbone hydrogen bonding interactions is important to formulating a more complete, molecular-level understanding of polyQ fibril formation mechanisms.

6.2 Results and Discussion

Infrared and Raman spectroscopies can specifically probe interactions between different molecular species. One approach is to correlate the solvatochromatic frequency shifts of specific vibrational bands in solvents of different polarities to variations in interaction energies and hydrogen bonding strengths,[146] as first shown by Badger and Bauer.[147] Recent work by the Boxer group[148,149] and others,[150,151] however, have shown that the solvatochromatic shifts of many probes can often, but not always, be attributed to the vibrational Stark effect. Consequently, frequency shifts due to the vibrational Stark effect can be used to determine the local electric fields felt by the probes.

Using UV Resonance Raman (UVRR) spectroscopy, we previously showed that vibrational modes localized on the amide groups of Q's side chains are sensitive to their local structure,[71] hydrogen bonding, and dielectric environments.[72] For example, the Amide I band of primary amides (denoted as AmI^P) sensitively probe the local hydrogen bonding of the Q side chain carbonyl group. The carbonyl stretching AmI^P band frequency and Raman cross section dramatically decrease in water relative to acetonitrile. Both of these spectral trends can be rationalized by the fact that, compared to acetonitrile, water stabilizes the primary amide ground state $-O-C=NH_2^+$ resonance structure compared to the $O=C-NH_2$ resonance structure.[71]

6.2.1 Relation Between Gln Side Chain AmI Frequency and ΔH_{int}

The well-studied Amide I bands of secondary amides (denoted as AmI^S) similarly show sensitivities to the solvation and local hydrogen bonding of the peptide bond backbone carbonyl groups.[89–93] Wang et al.[93] previously showed that the AmI^S frequency is linearly dependent on the solvent's acceptor number (AN), which is a measure of the strength of the solvent's hydrogen bonding, Lewis acid, and van der Waals interactions with solutes.[152] Using the AmI^S frequency-AN correlation, Wang et al. showed that the AmI^S frequency can be used to estimate the change in enthalpy due to these interactions (ΔH_{int}) for the secondary amide peptide bond carbonyl groups relative to that of these carbonyl groups in vacuum.

Inspired by Wang et al.'s[93] work, we investigated whether the AmI^P frequency could be used to determine the ΔHint of the primary amide carbonyl groups with their chemical environments. We initially focused on the small primary amide model compound formamide, because its ΔH_{int} values in various solvents are known.[153] We used the formamide solvent dependence of the AmI^P frequencies measured by Cutmore and Hallam[154] and the AmI^P frequency of formamide in water by Eaton et al.[155] to determine the frequency dependence on the solvent AN and donor number (DN).

Figure 6.1a shows the formamide AmI^P frequency dependence on the solvent AN and DN. There is a robust linear correlation between the AmI^P frequency and solvent AN. The least-squares linear fit obtained from the AmI^P frequency dependence on the solvent AN is:

$$AmI^P = b + m(AN) = 1715 cm^{-1} + (-0.50 cm^{-1} acc\#^{-1})(AN) \qquad \textbf{6.1}$$

where b=1715 cm^{-1} and m=-0.50 cm^{-1} acc $\#^{-1}$ for formamide.

The donor number is a measure of a solvent's hydrogen bond accepting strength and Lewis basicity, which do not affect solvent-amide carbonyl group interactions.[93] Therefore, as expected, there is no correlation between the AmI^{P} frequency and solvent DN.

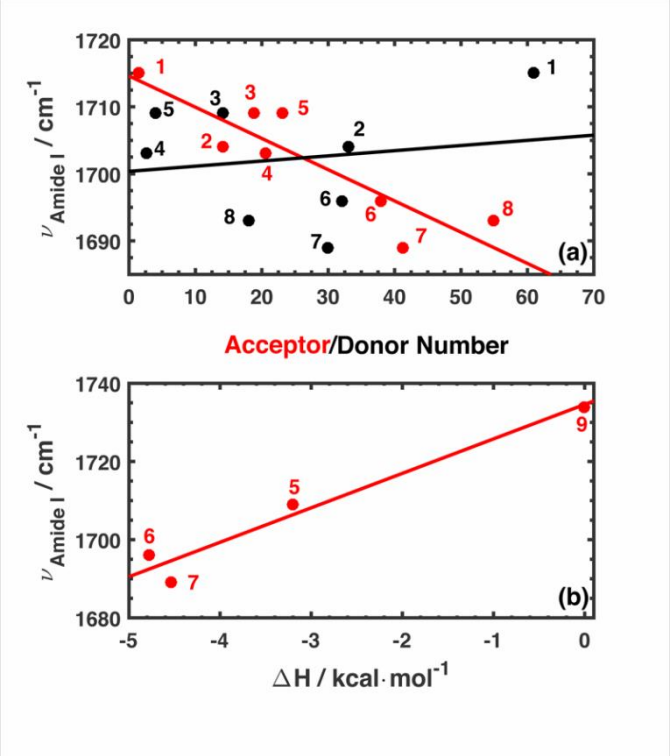

Figure 6.1: AmI^{P} frequency dependence of formamide on (a) solvent acceptor number, solvent donor number, and (b) interaction enthalpy. The number labels in the figure correspond to the following solvents: 1. triethylamine; 2. pyridine; 3. acetonitrile; 4. nitromethane; 5. chloroform; 6. ethanol; 7. methanol; 8. H_2O; 9. Vapor phase. The frequency dependence of formamide on the different solvents was obtained from data of Cutmore and Hallam[154] and Eaton et al.[33] Reprinted with permission from Punihaole et al. *J. Phys. Chem. Lett.* 2018, 9, 8 1944-1950. Copyright 2018 American Chemical Society.

We also examined the AmI^P frequency dependence on the ΔHint between the formamide carbonyl and different solvents using previously reported thermodynamic data.[153] For example, the ΔH_{int} values for the formamide carbonyl group in hydrogen bonding donor solvents such as ethanol, methanol, chloroform, and in the vapor phase are −4.8, −4.5, −3.2, and 0 kcal · mol^{-1}, respectively (Figure 6.1b), indicating that the AmI^P frequency linearly depends on ΔH_{int}, with m'=8.8 cm^{-1} $kcal^{-1}$ mol and b'=1735 cm^{-1} for formamide.

$$AmI^P = b' + m'(\Delta H_{int}) = 1735\ cm^{-1} + (8.8\ cm^{-1}kcal^{-1}mol)(\Delta H_{int}) \qquad \textbf{6.2}$$

Numerous studies have characterized the dependence of the AmI^P frequencies on the solvent AN. Unfortunately, few studies have characterized the dependence of the AmI^P frequencies on ΔH_{int}. Fortunately, we can convert the AmI^P frequency dependence upon AN to its dependence upon ΔHint by relating eqs.

6.1 and 6.2, as prescribed by Wang et al.[93]

First, the AN value of the solvent "vacuum" is determined by equating eqs.

6.1 and 6.2, for formamide in vacuum. We calculate that AN = −40 (acc #) for the "vacuum solvent". Thus, the eq. 6.2 intercept is b'=b +m (-40 (acc #)). The AN of vacuum, as well as that of other solvents, is independent of the solute present. Thus, the intercept of eq. 6.2 can be determined for any primary amide as long as the AmI^P frequency dependence on AN is known.

The AmI^P frequency is linearly proportional to both AN and ΔH_{int}. Thus, we can calculate the conversion factor between AN and ΔH_{int} for formamide from the slopes of eqs.

6.1 and 6.2:

$$m = \frac{-0.50\ cm^{-1}}{1\ acc\#} = m' = \frac{8.8\ cm^{-1}}{1\ kcal\ mol^{-1}}$$

$$(1\ acc\#^{-1}) = \frac{8.8\ cm^{-1} kcal^{-1} mol}{-0.50\ cm^{-1}} = -18\ kcal^{-1}\ mol \qquad \textbf{6.3}$$

Assuming that the conversion factor is constant for all primary amide compounds, eq. 6.2 can be written as follows:

$$AmI^P = b' + m'(\Delta H_{int}) \qquad \textbf{6.4}$$

$$= \big(b + m(-40\ acc\#)\big) + \big(-18\ kcal^{-1}\ mol\ m(\Delta H_{int})\big)$$

Using eq. 6.4, we can estimate ΔH_{int} for any primary amide compound from its AmI^P frequency dependence on AN.

6.2.2 Determining Side Chain ΔH_{int} in Polyglutamine Peptides

Because Q and polyQ peptides have very limited solubility in low AN solvents, we often utilize propanamide as an alternative compound to model the Q side chains.[71,72,82] We previously found, using UVRR spectroscopy, that the AmI^P frequency of propanamide ($CH_3CH_2CONH_2$) is 1669 cm^{-1} in water and 1692 cm^{-1} in acetonitrile.[72] Because the AN of water is 54.8, while that of acetonitrile is 18.9[156], we can calculate the linear dependence of the propanamide AmI^P frequency on AN. We find that m = −0.64 cm-1 $(acc\ \#)^{-1}$ and b= 1704 cm^{-1} for propanamide. By

substituting these values into eq. 6.4, we derive eq. 6.5 which can be used to estimate ΔH_{int} for propanamide.

$$AmI^{P}(cm^{-1}) = \ 1730\ (cm^{-1}) + (12\ cm^{-1}\ kcal^{-1}\ mol)(\Delta H_{int}) \qquad \textbf{6.5}$$

Because propanamide models the Q side chain, we can use eq. 6.5 to estimate the ΔH_{int} of side chain carbonyl groups in monomeric Q and in the polyQ peptide $D_2Q_{10}K_2$ (Q10) in any environment. The ΔH_{int} of each system is measured with respect to the vapor phase, where ΔH_{int} is zero. In the case of monomeric Q in aqueous solution, where the side chain carbonyl groups are hydrogen bonded to water, the AmI^P band is located at 1679 cm^{-1}.[82] As shown in Table 6.1, this frequency corresponds to a ΔH_{int} of −4.3 kcal · mol^{-1}.

Table 6.1: Estimated ΔH_{int} Values for Q Side Chain Carbonyl Groups in Different Peptide Conformations. Reprinted with permission from Punihaole et al. *J. Phys. Chem. Lett.* 2018, 9, 8 1944-1950. Copyright 2018 American Chemical Society.

	Am I^P (cm^{-1})	Expt. ($kcal \cdot mol^{-1}$)	MD ($kcal \cdot mol^{-1}$)	H-bonding Type
Q amino acid	1679	-4.3	–	s.c.-w
PPII	1680	-4.2	-4.4	s.c.-w
β-strand	1660, 1679	-5.8, -4.3	-4.7, -4.4, -4.6	s.c.-p.b.,s.c.-w,
Fibrils prepared from β-strand	1659	-5.9	-6.2	s.c.-s.c.
Fibrils prepared from PPII	1665	-5.4	-6.2	s.c.-s.c.

[a] s.c.: side chain; p.b.: peptide backbone; w: water

We recently showed that Q10 adopts two stable monomeric conformations in aqueous solution.[82] One solution conformation is a polyproline II (PPII) like structure, which can be prepared using a standard "disaggregation" protocol.[31] The side chain and peptide backbone amides of this PPII-like conformation are hydrogen bonded to water molecules. The other conformation is a collapsed β-strand-like conformation, which is simply prepared by dissolving the synthesized peptide in water. The side chains of the β-strand-like conformations are intramolecularly hydrogen bonded, as well as to solvating water molecules. Both PPII-like and collapsed β-strand-like Q10 conformations can be poised to aggregate into different amyloid-like fibril polymorphs when incubated in aqueous solutions at elevated temperatures for about one week. The cores of fibrils grown from both monomeric conformations are composed almost exclusively of antiparallel β-sheets, though slight structural differences between the two polymorphs are observed.[46] Both fibril polymorphs, however, form inter-amide side chain hydrogen bonds between neighboring strands within a given β-sheet.[46]

Our previous studies[46,82] examined the AmI^P frequencies of these Q10 fibrils, as well as of the PPII-like and β-strand-like solution monomers. Table 6.1 shows these frequencies, and

their corresponding estimated ΔH_{int} values determined using eq. 6.5. As expected, the ΔH_{int} of side chain carbonyl-water interactions in PPII-like and β-strand-like monomers are identical to those of monomeric glutamine in aqueous solution. In contrast, the ΔH_{int} of side chain carbonyl-peptide interactions that occur in the β-strand-like monomers are −5.8 kcal · mol^{-1}, a value similar to that estimated for side chain-side chain interactions in polyQ fibrils. For fibrils prepared from PPII-like monomers, we determine that the ΔH_{int} value of side chain-side chain interactions is −5.4 kcal · mol^{-1} while that prepared from β-strand-like monomers is found to be -5.9 kcal mol^{-1}. This result may suggest that fibrils prepared from the PPII monomer have weaker side chain-peptide interactions. However, the difference between ΔH_{int} values for fibrils grown from the two different monomer states is within the estimated error, as discussed in Appendix B.

In determining these ΔH_{int} values, we implicitly assume that there is no significant coupling between AmI^P oscillators. Strong coupling of neighboring oscillators would cause "excitonic" splitting of the AmI^P band frequency, which would complicate the determination of ΔH_{int}. In dilute solutions of amides, such as for monomeric glutamine in water, no coupling of the AmI^P vibrations can occur. In the case of Q10 PPII-like and β-strand-like peptide monomers, we expect that coupling will likely be weak, so that the impact of band splitting on the AmI^P frequency is essentially negligible.

In contrast, neighboring oscillator coupling could impact the AmI^P band of polyQ fibrils. In antiparallel β-sheet conformations, for example, the backbone AmI^S frequency is impacted by coupling with neighboring oscillators, which results in band splitting (vide infra). However, in the case of the AmI^P, we see no evidence of coupling between the neighboring oscillators of Q10 fibrils. Our previously reported UVRR AmI^P bands of Q10 fibrils are very narrow and consist of only single bands,[46] suggesting no excitonic splitting. This conclusion is further reinforced by the

fact that the Raman AmI^P frequencies of Q10 fibrils are essentially the same as reported IR frequencies of fibrils prepared from similar polyQ peptides.[48,157] This fact precludes the possibility that we only observe "bright" (Raman active) exciton modes of the AmI^P, but no "dark" (Raman inactive) modes.

6.2.3 Determining ΔH_{int} of Backbone-backbone Interactions in PolyQ Fibrils

We sought to compare the ΔH_{int} of side chain-side chain and backbone- backbone hydrogen bonds in Q10 fibrils. To estimate the ΔH_{int} of backbone-backbone interactions, we used an equation derived by Wang et al.[93] for the AmI^S band. However, the fibril AmI^S bands are impacted by coupling. As a result, it was first necessary to determine the uncoupled AmI^S frequencies of Q10 fibrils prepared from PPII-like and β-strand-like monomer solutions.

The excitonic splitting pattern of the AmI^S band in antiparallel β-sheets is well understood. Given the approximate D_2 symmetry of antiparallel β-sheets, the AmI^S vibration is predicted to split into four vibrational states, the A, B_1, B_2, and B_3 modes, all of which are Raman active. We curve fit the AmI^S bands of Q10 fibrils from our previously published data (Figure B.1) and assigned the A, B_1, B_2, and B_3 modes using the work of Krimm and coworkers[35–37] as a guide. Our spectral analysis of Q10 fibrils prepared from β-strand-like monomer solutions indicates that the A mode is located at 1665 cm^{-1}, the B_1 mode is at 1695 cm^{-1}, the B_2 mode is at 1625 cm^{-1}, and the B_3 mode is at 1680 cm^{-1}.[82] In contrast, for fibrils prepared from PPII-like monomers, the AmI^S A, B_1, B_2, and B_3 modes are located at 1660 cm^{-1}, 1688 cm^{-1}, 1617 cm^{-1}, and 1675 cm^{-1}, respectively.[46]

To estimate the unperturbed AmI^S frequencies, we utilize the perturbation theory approach developed by Miyazawa,[161] where the observed frequencies of the AmI^S bands are given by:

$$AmI^s(\delta,\delta') = v_0 + \sum_{s,t} D_{s,t}\cos(s\delta)\cos(t\delta') \qquad \textbf{6.6}$$

where ν_0 is the unperturbed AmI^S frequency, and $D_{s,t}$ are the interaction constants between peptide backbone amides separated by s amide groups and t chains along the t^{th} peptide chain. δ and δ' are the phase angles between adjacent AmI^S oscillators along a given peptide chain or between hydrogen bonded peptide backbone amides on neighboring chains.

Moore and Krimm[158–160] have shown that, for an infinite size antiparallel β-sheet system (which is an appropriate limit for Q10 fibrils), eq. 6.6 can be written as:

$$AmI^S(\delta,\delta') = v_0 + D_{00} + D_{10}\cos(s\delta) + D_{01}\cos(t\delta') + D_{11}\cos(s\delta)\cos(t\delta') \qquad \textbf{6.7}$$

For antiparallel β-sheets, the following set of equations can be written from eq. 6.7 by assuming different combinations of 0 or π phase angles:

$$\begin{aligned} AmI^S(0,0) &= v_0 + D_{00} + D_{10} + D_{01} + D_{11} \\ AmI^S(0,\pi) &= v_0 + D_{00} + D_{10} - D_{01} - D_{11} \\ AmI^S(\pi,0) &= v_0 + D_{00} - D_{10} + D_{01} - D_{11} \\ AmI^S(\pi,\pi) &= v_0 + D_{00} - D_{10} - D_{01} + D_{11} \end{aligned} \qquad \textbf{6.8}$$

where $AmI^S(0, 0)$ corresponds to the A mode, while $AmI^S(0, \pi)$, $AmI^S(\pi, 0)$, and $AmI^S(\pi, \pi)$ correspond to the B_1, B_2, and B_3 modes, respectively. The unperturbed frequency of the AmI^S cannot be exactly determined in eq. 6.7 since there are too many unknown coefficients to solve. However, we can estimate an effective unperturbed frequency by substituting ν_0 with:

$$\nu'_0 = \nu_0 + D_{00} \qquad \textbf{6.9}$$

Thus, using the experimentally observed AmI^S A, B_1, B_2, and B_3 mode frequencies for Q10 fibrils prepared from β-strand-like (PPII-like) monomer solutions, we determine ν' to be 1663 cm^{-1} (1660 cm^{-1}), D_{10} to be 6.3 cm^{-1} (7.5 cm^{-1}), D_{01} to be −21.3 cm^{-1} (−21.5 cm^{-1}), and D_{11} to be 13.8 cm^{-1} (14 cm^{-1}).

Our calculated ν', D_{01}, D_{10}, and D_{11} constants match in sign and are close in value to those determined by Krimm and coworkers for polyalanine and polyglycine antiparallel β-sheets[158–160]. This gives us confidence that ν' can be used to robustly estimate ΔH_{int} for inter-amide hydrogen bonds formed between peptide bonds in the core of polyQ fibrils.

To estimate ΔH_{int} from ν', we used the following equation derived by Wang et al.[93] for N-acetyltrialanine methyl ester:

$$AmI^S = 1699.3\ cm^{-1} + (9.46\ mol\ cm^{-1} kcal^{-1})(\Delta H_{int}) \qquad \textbf{6.10}$$

Using eq. 6.10, we estimate that the ΔH_{int} of backbone-backbone interactions are −3.8 kcal mol^{-1} for fibrils prepared from both Q10 β-strand-like and PPII-like monomers. This interaction strength is similar to NMA-NMA hydrogen bonding strengths previously measured.[162–164]

The above ΔH_{int} values indicate that backbone-backbone interactions in Q10 fibrils are enthalpically less favorable than side chain-side chain hydrogen bonding interactions. It is possible that the difference in the ΔH_{int} values for inter-amide backbone versus side chain interactions results from our use of ν'_0 instead of ν_0 when calculating ΔH_{int}. Krimm and coworkers estimate that D_{00} is relatively small (roughly −5 cm^{-1}). Using a D_{00} value of −5 cm^{-1}, we roughly estimate that side chain-side chain interactions are still enthalpically more favorable (by ∼−1 kcal · mol^{-1}) than backbone-backbone interactions.

6.2.4 Comparison to MD Simulations

In our previous work, we used metadynamics and molecular dynamics (MD) simulations, in conjunction with UVRR, to investigate the monomer solution-state and fibril structures of Q10.[46,82] Here, we use our previously simulated Q10 structures as initial coordinates to perform classical MD simulations on solution-state and fibril Q10 conformations. We use these simulations to independently calculate ΔH_{int} values of various backbone and side chains interactions and compare them to that obtained via UVRR.

Figure 6.2 shows the simulated PPII-like, β-strand-like, and fibril structures. For each structure, we determined the average number of inter-amide (e.g. side chain-side chain, side chain-backbone, and backbone-backbone) and amide-water hydrogen bonding interactions

formed per peptide (Figure 6.2). We define a hydrogen bond as having a geometry such that the heavy atom donor to acceptor distance is <3.0 Å with a bond angle of 180° (±30).

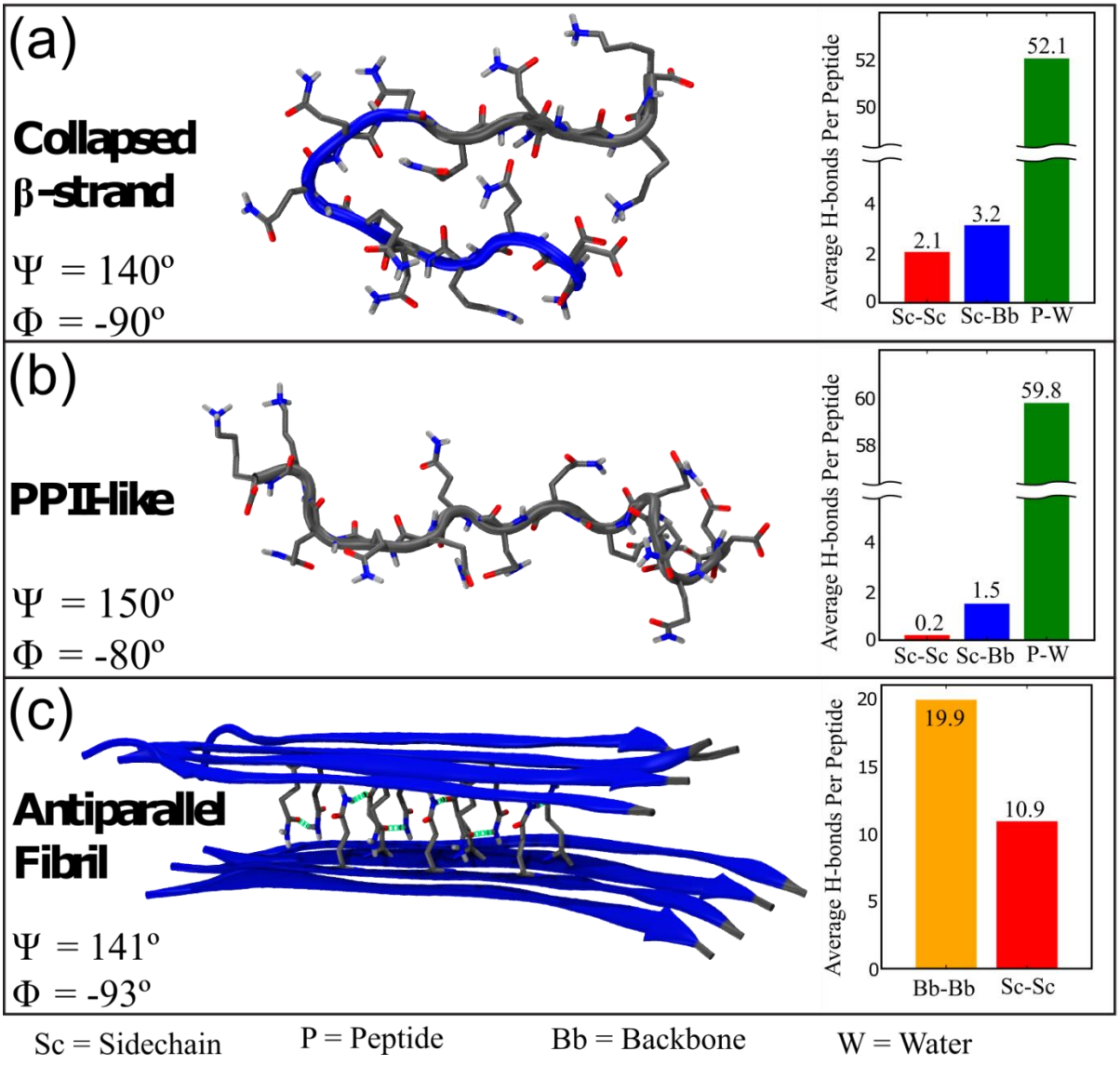

Figure 6.2: Structures and backbone Ramachandran angles of MD simulated Q10 structures in the (a) monomeric collapsed β-strand-like, (b) monomeric PPII-like, and (c) antiparallel fibril conformations. The bar graphs show the average number of various hydrogen bonding interactions per peptide residue. Reprinted with permission from Punihaole et al. *J. Phys. Chem. Lett.* 2018, 9, 8 1944-1950. Copyright 2018 American Chemical Society.

The number of backbone-backbone hydrogen bonds is negligible for both the PPII-like and β-strand-like monomeric Q10 structures. For the monomeric PPII-like structure, we find that 94 % of side chain amides are hydrogen bonded to water, 2 % of side chains are hydrogen bonded to backbone amides, and 4 % of side chains are hydrogen bonded to other side chains. For the β-strand-like structure, 66 % of side chain amides are hydrogen bonded to water, 20 % of side chains are hydrogen bonded to backbone amides, and 16 % of side chains are hydrogen bonded to each other. In contrast, the Q10 antiparallel fibril structure's interior backbone and side chain amides are hydrogen bonded exclusively to other backbone and side chain amides.

From our MD simulations, we calculated average ΔH_{int} values for each simulated structure. We define ΔH_{int} as the sum of the Lennard-Jones (ΔE^{LJ}_{int}) and Coulombic potential energy (ΔE^{elec}_{int}) terms, where the Δ signifies the difference in energy of the carbonyl oxygen atoms with their interacting partner atoms in close proximity relative to being infinitely separated. Due to the negligible change in volume throughout the simulations, the ΔH_{int} can be accurately approximated as the change in the energy of interaction (ΔE_{int}):

$$\Delta H_{int} \approx \Delta E_{int} = \Delta E_{int}^{elec} + \Delta E_{int}^{LJ} \qquad \textbf{6.11}$$

When calculating ΔH_{int}, we define interacting groups as those with heavy atoms at distances of less than or equal to 5 Å. Thus, our calculated ΔH_{int} is not limited to strong hydrogen bonding interactions, which generally occur for heavy atom distances less than 3 Å. For the Q10 antiparallel β-sheet conformation (Figure 6.2c), the six-innermost buried side chain and backbone amide groups were used to calculate the ΔH_{int}, since they best model the interior of the fibril core.

The ΔH_{int} values calculated from the MD simulated Q10 structures are shown in Table 6.1. Overall, these calculated ΔH_{int} values from the MD simulations are in excellent agreement with our experimentally determined values. We calculate an average ΔH_{int} value of $-6.20(\pm 0.63)$ kcal $\cdot$ mol^{-1} for side chain-side chain interactions and $-4.30(\pm 0.67)$ kcal $\cdot$ mol^{-1} for backbone-backbone interactions for the Q10 fibril model. For both the PPII-like and the collapsed β-strand-like monomer structures, we determine that the ΔH_{int} of side chain carbonyl-water interactions is $-4.4(\pm 0.61)$ kcal $\cdot$ mol^{-1}. For the β-strand-like monomer structure, we calculate that the ΔH_{int} of side chain-backbone hydrogen bonding is $-4.70(\pm 0.57)$ kcal $\cdot$ mol^{-1} while that of side chain-side chain hydrogen bonding is $-4.63(\pm 0.53)$ kcal $\cdot$ mol^{-1}.

6.3 Conclusion

Our experimental and computational results indicate that the ΔH_{int} of side chain-side chain and side chain-backbone interactions of Q10 fibril and monomeric β-strand-like structures are enthalpically more favorable than side chain-water. Interestingly, our results also indicate that, in polyQ fibrils, side chain-side chain interactions are more favorable than backbone-backbone interactions. The importance of this work is that, to our knowledge, these are the first results that experimentally quantify the relative enthalpic favorability of hydrogen bonding interactions in solution-state and fibril polyQ peptides. This work further validates the hypothesis that inter-amide side chain and backbone interactions play important roles in thermodynamically driving polyQ-rich proteins towards fibril structures.

6.4 Computational Details

MD simulations were performed using the NAMD software,[107] and VMD was used for visualization and analysis. All simulations were conducted under constant atom number, pressure (1 atm), and temperature (300 K) (NPT). The CHARMM36[109] force field was utilized for potential energy and force calculation. The initial coordinates of the simulated structures shown in Figure 6.2 were from our previously published, experimentally validated work.[46,82] Each system was solvated in TIP3P water.[165] The simulated model fibril system, including water, consisted of a total of 20,284 atoms. The simulations for monomeric PPII-like and β-strand-like Q10 structures with explicit water each consisted of 17,189 atoms. Each system was simulated for 5 ns, and 500 snapshots were extracted from the simulations for analysis. The initial coordinates of each simulated structure, as well as the scripts used in our analysis, are available for download.

6.5 Acknowledgements

Funding for this work was provided by the University of Pittsburgh (DP, RSJ, SAA), the Defense Threat Reduction Agency HDTRA1-09-14-FRCWMD (RSJ, SAA) and NIH R01 DA027806 (RJW). MD simulation computer time was supported by XSEDE MCB060069 and computer equipment was purchased from NSF funds (CHE-1126465 and P116Z080180). We thank the late Prof. Jeffry Madura (Duquesne University, Department of Chemistry and Biochemistry) for useful discussions. Also, we thank Stephen White (University of Pittsburgh, Molecular Biophysics and Structural Biology Program), Tim Coleman (University of Pittsburgh,

Department of Statistics), and Prof. Satish Iyengar (University of Pittsburgh, Department of Statistics) for assistance in our error estimation.

6.6 Supporting Information

Supporting information can be found in Appendix B and includes the following: fitting of AmI^S UVRR spectra of Q10 fibrils and estimation of error for determining ΔH_{int} values.

7.0 UV Resonance Raman Structural Characterization of an (In)Soluble Polyglutamine Peptide

Adapted with permission from: Ryan S. Jakubek, Stephen E. White, and Sanford A. Asher. "UV Resonance Raman Structural Characterization of an (In)soluble Polyglutamine Peptide." The journal of physical chemistry B, 2019, 123, 8, 1749-1763.

Author Contributions: RSJ collected, analyzed, and interpreted the Raman and absorbance data with the assistance of SEW. SEW collected the TEM images of Q20 fibrils. The manuscript was prepared by RSJ and SAA with assistance by SEW.

Fibrillization of polyglutamine (polyQ) tracts in proteins is implicated in at least ten neurodegenerative diseases. This generates great interest in the structure and the aggregation mechanism(s) of polyQ peptides. The fibrillization of polyQ is thought to result from the peptide's insolubility in aqueous solutions; longer polyQ tracts show decreased aqueous solution solubility, which is thought to lead to faster fibrillization kinetics. However, few studies have characterized the structure(s) of polyQ peptides with low solubility. In the work here we use UV resonance Raman spectroscopy to examine the secondary structures, backbone hydrogen bonding, and side chain hydrogen bonding for a variety of solution state, solid, and fibril forms of $D_2Q_{20}K_2$ (Q20). Q20 is insoluble in water and has a β-strand-like conformation with extensive inter- and intra-peptide hydrogen bonding in both dry and aqueous environments. We find that Q20 has weaker backbone-backbone and backbone-side chain hydrogen bonding and is less ordered compared to that of polyQ fibrils. Interestingly, we find that the insoluble Q20 will form fibrils when incubated in water at room temperature for ~5 hours. Also, Q20 can be prepared using a well-known disaggregation procedure to produce a water soluble PPII-like conformation with negligible inter- and intrapeptide hydrogen bonding and a resistance to aggregation.

7.1 Introduction

Expanded polyglutamine (polyQ) tracts in proteins and peptides induce aggregation and fibrillization.[1] This polyQ-induced fibrillization is associated with at least ten neurodegenerative diseases, including Huntington's disease (HD) and multiple spinocerebellar ataxias (SCAs).[1] The mechanism of toxicity and the identity of the toxic species are still debated.[14,16,166] The common factor for polyQ-associated neurodegenerative diseases is the presence of an expanded polyQ tract.[1]

In polyQ repeat diseases, longer polyQ tracts are correlated with an earlier disease symptom age-of-onset.[167] Disease symptoms are only observed when the protein polyQ tract length surpasses a critical length (~≥36Q for the huntingtin protein in HD).[1] This polyQ tract length dependence of disease age-of-onset is thought to result from a length-induced increase in the polyQ aggregation kinetics.[47,168,169]

This possibility was strengthened when Chen et al. showed that polyQ peptides with longer polyQ tracts have faster aggregation kinetics.[47] Also, Chen et al. used aggregation rates calculated from polyQ peptides at high concentrations (~5-50 μM) to extrapolate aggregation rates for polyQ peptides at physiological concentrations (~0.1 nM).[47] In these calculations, Q47 at physiological concentrations was calculated to aggregate in ~31 years. This aggregation rate is quite similar to the HD age-of-onset (30-40 years) for patients with a Q47 tract length in the huntingtin protein. Also, Q36 and Q28 lengths were calculated to begin aggregating in 141 and 1273 years, respectively, at physiological concentrations. These putative ages of disease onset

roughly agree with the age of onset for polyQ tracts with <36Q residues in the huntingtin protein that do not produce HD symptoms within a patient's lifetime. Because of the associations between polyQ tract length, aggregation kinetics, and disease age-of-onset, there is much interest in understanding the aggregation mechanism(s) of polyQ peptides of different lengths. Unfortunately, these studies of polyQ peptides are limited by the accompanying low solubility of these polyQ peptides in water.

For example, polyQ peptides with ~≥20 Q residues generally show low water solubility. To study the polyQ solution-state structure and fibrillization kinetics of long polyQ peptides, Chen et al. developed a "disaggregation" procedure that increases polyQ peptide solubility.[31] Their solubilization procedure is referred to as disaggregation because it is thought to remove trace aggregate oligomers that are believed to seed fibrillization.[33] Unfortunately, the exact mechanism by which this disaggregation protocol solubilizes polyQ peptide monomers is poorly understood. Because of the low solubility of long polyQ peptides, most studies use polyQ peptide solutions that are disaggregated before study.

Recently, Punihaole et al.[82] used UV resonance Raman (UVRR) spectroscopy and metadynamics simulations to examine the solution-state structure of the small polyQ peptide $D_2Q_{10}K_2$ (Q10) in its disaggregated (DQ10) and non-disaggregated (NDQ10) forms. They found that aqueous NDQ10 has a collapsed β-strand-like conformation with significant intra-peptide hydrogen bonding. In contrast, DQ10 has a Polyproline II- (PPII-) like conformation with negligible inter- and intra-peptide hydrogen bonding. They showed that the β-strand-like (NDQ10) and the PPII-like (DQ10) structures are both predominantly monomeric with large activation energy barriers between them that prevent interconversion between these solution-

state monomeric structures. They also showed that NDQ10 and DQ10 are both soluble in water at concentrations of up to 1 mg/mL.

Computational [37,38] and experimental [43,44] studies have shown that longer polyQ peptides increasingly possess peptide-peptide hydrogen bonding. Studies showed that water is a good solvent for Q<16, a theta solvent for Q=~16, and a poor solvent for Q>16.[43–45] Apparently, the increasing favorability of interpeptide interactions for longer polyQ peptides results in decreased aqueous solubilities that are thought to drive the formation of peptide aggregates and fibrils.[43,45] The peptide-peptide interactions of dilute polyQ solutions in poor solvents are satisfied via intra-peptide hydrogen bonds that give rise to compact collapsed structures.[38,43,45] In contrast, more concentrated polyQ solutions will form inter-peptide interactions that result in polyQ peptide aggregation.[38,43,45] The low solubility of longer polyQ peptides is thought to promote their aggregation.[38,43] However, little is known about the structures and fibrillization of polyQ peptides with low aqueous solubility.

Here, we use UVRR spectroscopy to examine the structures of $D_2Q_{20}K_2$ (Q20). We find that non-disaggregated Q20 (NDQ20) is insoluble in water. In this work we describe a polyQ peptide as insoluble if the peptide forms a pellet upon centrifugation (21,130 x g for 30 min).

Figure 7.1 summarizes the structures of Q20 observed in this study. Q20 was synthesized by Thermo Fisher Scientific using Fmoc-based solid-phase peptide synthesis (SPPS) (Figure 7.1a). The peptide was purified in 0.05% TFA using reverse-phase high-pressure liquid chromatography (HPLC) (Figure 7.1b). The purified peptide was then lyophilized, yielding the solid-phase peptide synthesis Q20 (SPPS Q20) species, which occurs in the form of a white powder (Figure 7.1c). For all experimental work, SPPS Q20 is considered to be the initial state

of the peptide. Using UVRR we find that SPPS Q20 is in a β-strand-like conformation (Figure 7.1c).

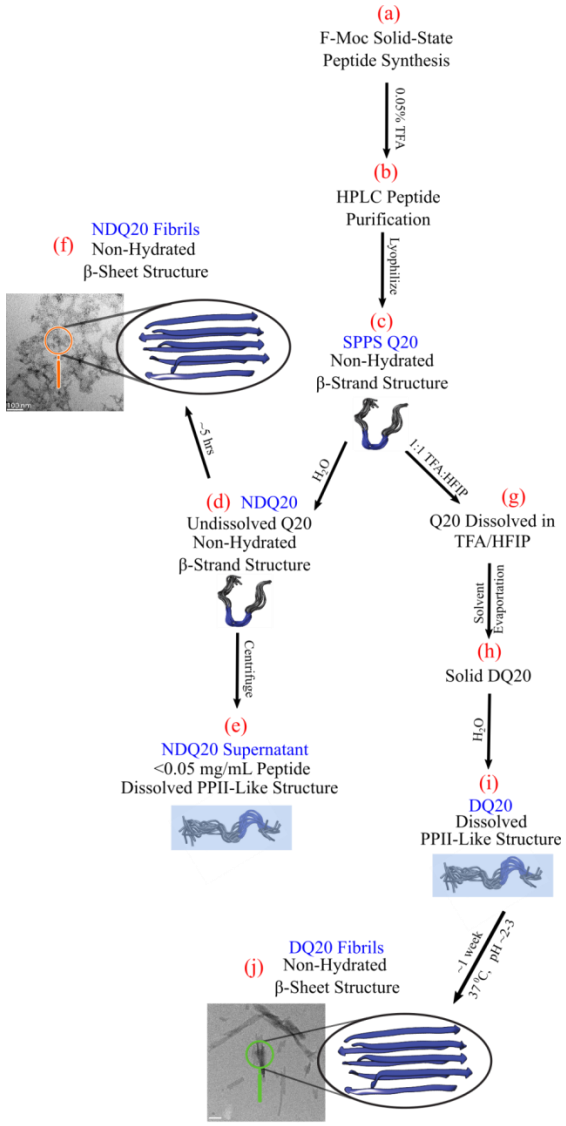

Figure 7.1: Summary of the forms of Q20 examined. The blue text indicates the nomenclature for this particular form of Q20. The peptide models shown are from simulations previously calculated for Q10 by Punihaole et al.[46,82] They depict the peptide structure of each state examined. The structures shaded in blue are found to have the backbone and Gln side chain primary amides hydrogen bonded to water, while structures not shaded in blue have amide groups that are not hydrogen bonded to water. These structures are discussed in detail within the text. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

SPPS Q20 added to water does not form a clear solution at 0.3 mg/mL. This state of Q20 is designated as non-disaggregated Q20 (NDQ20) because it was not disaggregated before being added to water (Figure 7.1d). We find that the secondary structure of NDQ20 is similar to that of the SPPS Q20. Despite the low water solubility of NDQ20, we find that NDQ20 forms β-sheet fibrils (Figure 7.1f) when incubated in water at room temperature.

We also prepared aqueous Q20 using the disaggregation protocol described in the methods section. SPPS Q20 will dissolve in a 1:1 mixture of trifluoroacetic acid (TFA) and hexafluoroisopropanol (HFIP) to form a clear solution (Figure 7.1g). The TFA/HFIP solvent can then be evaporated, and the resulting disaggregated Q20 (DQ20) peptide (Figure 7.1h) dissolves in water (Figure 7.1i). Using UVRR spectroscopy we find that the DQ20 peptide has a PPII-like secondary structure with backbone and side chain amide groups hydrogen bonded to water. DQ20 forms fibrils when incubated at 37°C and neutral pH for ~1 week (Figure 7.1j).

Finally, we find that after ultracentrifugation of NDQ20 a small amount of peptide remains in the supernatant (Figure 7.1e). The peptide in the NDQ20 supernatant has a PPII-like structure similar to that of the highly soluble DQ20.

7.2 Materials and Methods

7.2.1 Materials

$D_2Q_{15}K_2$ (Q15) (≥98% purity), $D_2Q_{20}K_2$ (Q20) (≥98% purity), and Trifluoroacetic acid (TFA) (≥ 99.5% purity) were purchased from Thermo Fisher Scientific. 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) (~99% purity) was purchased from Acros Organics.

7.2.2 Peptide Synthesis

Q15 and Q20 were synthesized by Thermo Fisher Scientific using Fmoc-based solid-phase peptide synthesis. The final peptide was cleaved from the solid support using TFA. The peptide was purified in 0.05% TFA using reverse-phase HPLC. The purified peptide was lyophilized to produce SPPS Q20 (Figure 7.1c).

7.2.3 Sample Preparation

NDQ20 (Figure 7.1d) was prepared by adding the SPPS Q20, as received from Thermo Fisher Scientific, to nanopure water. Samples were prepared in sterilized centrifuge tubes to prevent impurities from seeding aggregation.

The NDQ20 supernatant (Figure 7.1e) was made by first preparing NDQ20 at 1 mg/mL. The mixture was vortexed for ~5 minutes to dissolve any soluble peptide. After vortexing, the sample remained turbid. The sample was centrifuged for 30 minutes at 21,130 x g and then for 30 min at 355,524 x g. The top ~50% of the supernatant was used for UV absorbance and UVRR measurements.

DQ20 (Figure 7.1i) was prepared by using the disaggregation procedure developed by Chen et al.[31] To disaggregate polyQ, the peptide was incubated in a 1:1 TFA and HFIP mixture for ~2-4 hours (Figure 7.1g). The solvent was then evaporated with a stream of dry nitrogen gas (Figure 7.1h). The peptide was then dissolved in water to a concentration of 0.3 mg/mL.

TFA samples at pH=~+0.5 were prepared by adding 10% (v/v) TFA to nanopure water. TFA samples (10% (v/v)) in acidic conditions, pH=~-1.5, were prepared by adding TFA to concentrated hydrochloric acid. TFA in basic conditions was prepared by adding a known

volume of 10 M NaOH to a 10% (v/v) TFA solution to a final pH of ~12. The internal UVRR intensity standard sodium perchlorate ($NaClO_4$) was added to TFA samples and to the NDQ20 supernatant by adding a known volume of 5M $NaClO_4$ to the sample.

7.2.4 Absorbance Measurements

Absorbance measurements used a Cary 5000 UV-Vis-NIR Spectrophotometer (Varian, Inc.) with a 0.02 cm pathlength, fused silica cylindrical cuvette.

7.2.5 UV Resonance Raman Instrumentation

The UVRR instrumentation was described in detail by Bykov et al.[103] Briefly, the third harmonic of an infinity Nd:YAG laser (Coherent) was Raman shifted (30 psi, H_2) to 204 nm (the 5^{th} anti-Stokes line of hydrogen). The Raman-scattered light was dispersed using a double monochromator in a subtractive configuration.[103] The spectrum was imaged using a liquid nitrogen-cooled CCD camera with a Lumagen E coating (Spec10:400B, Princeton Instruments). Samples were placed in a Suprasil fused silica NMR tube that was spun during the measurement to reduce sample photodegradation. A ~165° backscattering geometry was used to collect the Raman scattering.

7.2.6 Transmission Electron Microscopy (TEM)

TEM images of DQ20 and NDQ20 fibrils were collected using a Morgagni 268(D) electron microscope (FEI) at 89,000x and 140,000x magnification, respectively, using an

electron accelerating voltage of 80 kV. Images were recorded on a 10-megapixel ORCA camera (Hamamatsu). EM sample grids were prepared by incubating 3 μL of DQ20 or NDQ20 fibrils on carbon-coated copper EM grids for ~3 min, and excess solution was removed by blotting with filter paper. The grid was then stained with 3 μL of 2% (w/v) uranyl acetate solution for ~45 sec before blotting.

7.2.7 UV Resonance Raman Methods

UVRR excitation at ~204 nm is in resonance with the $\pi \rightarrow \pi^*$ electronic transitions of amide groups, which include the secondary amide peptide bond and primary amide glutamine (Gln) side chains.[62] This selectively enhances vibrational motions that couple to these electronic transitions. Thus, our UVRR spectra of polyQ peptides are dominated by vibrations localized on the backbone peptide bonds and the Gln side chain amide groups. This greatly simplifies the Raman spectra. UVRR spectra are sensitive to the structure and solvation states of the peptide.[63]

7.2.8 $AmIII_3^S$ Band Reports on the Backbone Ramachandran Ψ Angle

The peptide backbone amide III_3 band ($AmIII_3^S$) frequency was shown to be sinusoidally correlated to the peptide backbone Ramachandran Ψ angle.[64,65] Mikhonin et al. developed a method to calculate the Ψ angle from an experimentally measured $AmIII_3^S$ frequency for a peptide backbone in different solvation states and temperatures.[64] Asher et al. then showed that we can estimate the distribution of Ψ angles in a peptide by modeling the inhomogenously broadened $AmIII_3^S$ band as a sum of Lorentzian bands that approximate the homogenously broadened $AmIII_3^S$ bands.[170] These methodologies have been extensively used to examine the

secondary structure of peptides.[46,57,62,63,76,78–80,82] A detailed discussion of the equations used to calculate the Ψ angle distributions in this work can be found in Appendix C.

7.2.9 AmI Bands Report on the Hydrogen Bonding and Dielectric Environment of the Amide Carbonyls

The amide I (AmI) bands of the secondary amide peptide backbone (AmI^S) and the primary amide Gln side chains (AmI^P) predominantly involve C=O stretching. The AmI band frequency and intensity are sensitive to the dielectric constant and the hydrogen bonding of the amide carbonyl groups.[72,81,90–93,171] This makes the AmI band a spectral marker for examining the water exposure and hydrogen bonding of amide groups.[63,81] An environment with a large dielectric constant, such as water, increases the contribution of the amide dipolar resonance structure ($^-$O-C=NH$^+$) compared to that of the less-polar resonance structure (O=C-NH). This decreases the C=O bond force constant and the AmI frequency.[72,171] Also, stronger hydrogen bonding to the C=O bond decreases its bond force constant downshifting the AmI band frequency.[72,81,93]

The dielectric constant and the hydrogen bonding of the amide group also affect the AmI band UVRR intensities. In general, the deep UVRR enhancement of the AmI band of primary and secondary amides decreases with an increasing dielectric constant and/or increased hydrogen bonding to the amide C=O group.[72,171] This occurs because resonance enhancement depends on the Frank-Condon overlap between the amide ground and resonant excited states. The resonance enhancement of the AmI vibration scales with the square of the displacement between the equilibrium C=O bond ground electronic state and the ππ*excited state along the AmI vibrational normal coordinate.[172] The C=O bond in the ππ* excited state is typically

elongated compared to that in the ground state.[86,96] Thus, elongation of the C=O bond in a strong hydrogen bonding and/or high dielectric environment decreases the C=O bond displacement between the ππ* ground and excited state resulting in decreased UVRR enhancement. [72,171]

However, changes in the effective dielectric constant and hydrogen bonding also affect the amide ππ* excited state geometry.[72,86,91,92] Thus, a complete understanding of the dependence of the AmI UVRR intensity on environment requires knowledge on how the dielectric constant and hydrogen bonding of the amide group affects the C=O bond length of both the ground and ππ* excited state.[86] Table 7.1 summarizes the effects of changes in the dielectric environment and hydrogen bonding of the amide group on the AmI band intensity and frequency.

Table 7.1: Effects of Dielectric Constant (ε) and C=O Hydrogen Bonding on the Frequency and Intensity of the AmI UVRR Band. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

	ε Increase	ε Decrease	H-bond strength Increase	H-bond strength Decrease
Δ Intensity	Decrease	Increase	Decrease	Increase
Δ Frequency	Decrease	Increase	Decrease	Increase

7.2.10 UVRR Band Assignments of PolyQ Peptides

The UVRR spectra of solution-state[82] and fibril-state[46] Q10 were previously assigned in detail. The UVRR spectra of Q20 measured here are similar to those previously measured for Q10. Here, we briefly discuss the assignments of the conformationally sensitive UVRR bands. Please refer to our work on Q10 fibrils[46] and on Q10 in solution[82] for details.

7.2.11 $AmIII_3^S$ Band Assignments

The $AmIII_3^S$ band is found in the ~1200-1300 cm^{-1} spectral region. As discussed above, this band is sensitive to the Ψ Ramachandran angle of the peptide backbone. Punihaole et al.[82] previously found that DQ10 has $AmIII_3^S$ peaks at ~1275 cm^{-1}, ~1250 cm^{-1}, and ~1215 cm^{-1} that derive from Ψ angle populations centered at ~175°, ~150°, and ~10°, respectively. The ~150° Ψ angle distribution is characteristic of PPII-like secondary structures, while Ψ angle populations at 10° and 175 ° are characteristic of turn-like and 2.5_1-helix-like structures, respectively. From the UVRR data along with metadynamics simulations, Punihaole et al. concluded that DQ10

contains short PPII-like helices separated by turn regions.[82] The 2.5_1-helix-like conformation was found to be localized within the charged aspartic acid (Asp) and lysine (Lys) residues at the peptide N- and C-termini, respectively.

In contrast, NDQ10 has a single Raman $AmIII_3^S$ band at ~1240 cm^{-1}. This corresponds to a Ψ angle distribution centered at ~140°, which is characteristic of a β-strand-like conformation. From the $AmIII_3^S$ band as well as metadynamics simulations, Punihaole et al. concluded that NDQ10 occurs in a collapsed β-strand-like conformation.[82] A list of $AmIII_3^S$ band frequencies and secondary structure assignments for DQ10 and NDQ10 is shown in Table 7.2.

Table 7.2: $AmIII_3^S$ Band Assignments and Structures for NDQ10 and DQ10. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

	$AmIII_3^S$ Band Freq. (cm^{-1})	Ψ Angle (°)	Secondary Structure
$DQ10^{82}$	~1275	~175	2.5_1-Helix-Like
$DQ10^{82}$	~1250	~150	PPII-Like
$DQ10^{82}$	~1215	~10	Turn-Like
$NDQ10^{82}$	~1240	~140	β-Strand-Like

7.2.12 AmI Band Assignments

The AmI^S and AmI^P bands are both found in the ~1650-1700 cm^{-1} spectral region and spectrally overlap. Xiong et al.[57] showed that the AmI^P and AmI^S bands can be separately highlighted in the UVRR spectrum by collecting UVRR spectra at two different excitations wavelengths: 197 nm and 204 nm. Because the ππ* transition of primary amides is at higher energy than secondary amides, excitation deeper in the UV (197 nm) increases the resonance enhancement of the AmI^P band relative to the AmI^S band. By subtracting the 204 nm UVRR spectrum from that of 197 nm excitation (197-204 nm) we highlight the primary amide vibrations, including the AmI^P band. In addition, we can highlight the secondary amide vibrations, such as the AmI^S band, by subtracting the 197-204 nm difference spectrum from the 204 nm spectrum (204-(197-204) nm).

Punihaole et al. previously examined the AmI^P and AmI^S bands of NDQ10 and DQ10 peptides in their solution-[82] and fibril-state[46] conformations. The AmI^P bands of Gln side chains hydrogen bonded to water have low UVRR intensities and are broad, with center frequencies of ~1680 cm^{-1}. A low AmI UVRR intensity and a frequency of ~1680 cm^{-1} was previously observed for the AmI^P band of aqueous Gln[71] as well as for the water hydrogen bonded amides of Gln side chains of aqueous polyQ peptides.[82] Similarly, the AmI^S band of water hydrogen bonded secondary amides has a low UVRR intensity and a frequency of ~1700 cm^{-1}.[93]

Gln side chain and peptide backbone amide groups involved in peptide-peptide hydrogen bonding will be partially shielded from water. Thus, they will experience an environment with stronger hydrogen bonding and a lower dielectric constant compared to that of fully water exposed amide groups. For Gln side chain and peptide backbone amides involved in peptide-peptide hydrogen bonding, we experimentally observe intense AmI^P and AmI^S bands located at ~1660 cm^{-1}. This was previously observed for the AmI^P bands of Q10 in the solution-state[82] and in Q10 fibrils[46] as well as for the AmI^S bands of Q10 fibrils.[81]

Wang et al.[93] showed that the frequency of the AmI^S band of the peptide backbone is correlated to the enthalpy of interaction (ΔH_{int}) between the backbone amide C=O groups and their environment. The ΔH_{int} is dominated by strong interactions such as hydrogen bonds.[152] This allows for the calculation of the strength of a hydrogen bond between the backbone C=O groups and a hydrogen bond donor.[81,93] Punihaole et al.[81] later expanded this technique to examine the ΔH_{int} of Gln side chain C=O groups with their environment. They derived the following equation to estimate the interaction enthalpy of the Gln side chain:

$$\mathbf{AmI^{P}(cm^{-1}) = 1730\ (cm^{-1}) + (12\ cm^{-1}kcal^{-1}\ mol)(\Delta H_{int})} \qquad \mathbf{7.1}$$

Using eq. 7.1, Punihaole et al.[81] estimated the strengths of Gln side chain hydrogen bonding in solution-state polyQ peptides, and backbone and side chain hydrogen bonding in polyQ fibrils. They found that in both solution and fibril-state polyQ, side chain-side chain and side chain-backbone (~-5.9 kcal/mol) hydrogen bonding interactions are stronger than that of side chain-water (~-4.3 kcal/mol), backbone-water (~-4.3 kcal/mol), and backbone-backbone (~-3.8 kcal/mol) hydrogen bonding.[81]

It is important to note that calculating ΔH_{int} from an AmI frequency assumes that the AmI frequency is only dependent on ΔH_{int}. However, the AmI^{S} frequency of the peptide backbone is also strongly dependent on the peptide secondary structure.[173] This occurs predominantly through changes in the transition dipole coupling (TDC) of adjacent AmI^{S} oscillators.[173] As a result, the determination of peptide backbone ΔH_{int} is confounded by TDC. In contrast, TDC is weak and can be neglected for Gln side chain AmI^{S} oscillators because of the larger distances and decreased order between adjacent oscillators.[81] Punihaole et al. demonstrated this by showing that, in polyQ fibrils, the AmI^{S} band is split due to TDC while the AmI^{P} band consists of a single narrow peak that indicates negligible coupling.[81]

From the AmI frequency we can determine if the Gln side chain and backbone amide C=O groups are hydrogen bonded to water or peptide. However, the AmI^{P} and AmI^{S} bands have the same frequency (~1660 cm^{-1}) for C=O hydrogen bonding to backbone NH and Gln side chain NH hydrogen bond acceptors.[46,81] Therefore we cannot differentiate between backbone-side chain and backbone-backbone hydrogen bonding as well as side chain-backbone and side chain-side chain hydrogen bonding. In this paper side chain-peptide (backbone-peptide) hydrogen

bonding denotes that a side chain (backbone) carbonyl hydrogen bond acceptor is hydrogen bonded to either a side chain or backbone hydrogen bond donor. Table 7.3 summarizes the AmI band frequencies, intensities, and interaction enthalpies for polyQ peptide amide groups involved in different hydrogen bonding interactions.

Table 7.3: Characteristic AmI Frequencies and UVRR Intensities for Side Chain and Backbone Amide Groups Involved in Peptide-Peptide or Peptide-Water Hydrogen Bonding. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

Hydrogen Bonding Interaction	AmI Freq. (cm^{-1})	Relative AmI UVRR Intensity	Interaction Enthalpy (kcal/mol)	Sample(s)
Sidechain C=O-Water	~1680 (AmI^P)	Weak	~-4.3[81]	DQ10[81,82], NDQ10[81,82], Gln[71]
Sidechain C=O-Backbone NH	~1660 (AmI^P)	Strong	~-5.9[81]	NDQ10[81,82]
Sidechain C=O- Sidechain NH	~1660 (AmI^P)	Strong	~-5.9[81]	NDQ10[81,82], NDQ10 Fibrils[46,81], DQ10 Fibrils[46,81]
Backbone C=O-Water	~1700 (AmI^S)	Weak	~-4.3[93]	$NAcA_3ME$[93]
Backbone C=O- Backbone NH	~1660 (AmI^S)	Strong	~-3.8[81]	NDQ10 Fibrils[46,81], DQ10 Fibrils[46,81]

$NAcA_3ME$, N-aceyletrialanine methyl ester

7.3 Results and Discussion

7.3.1 Solubility of Q10, Q15, and Q20

We qualitatively examined the solubility of DQ15, NDQ15, DQ20, and NDQ20 by adding these peptides to water, vortexing, and examining the sample for separation upon standing or centrifugation at 21,130 x g for 30 minutes. We find that NDQ15, DQ15, and DQ20 all form apparently clear, homogenous solutions in water, as observed by eye, at 1 mg/mL concentrations. They do not form pellets upon centrifugation. Additionally, Punihaole et al. [82] previously showed that NDQ10 and DQ10 form clear solutions in water at 1 mg/mL concentrations, and they showed using diffusion-ordered NMR spectroscopy (DOSY) that these peptides are predominately monomeric. Also, using DOSY we find that DQ15 and DQ20 diffuse at a rate consistent with a monomeric peptide (data not shown). A detailed description of the DQ15 and DQ20 DOSY data will be presented in a future publication.

In contrast, NDQ20 does not form a clear, homogenous solution in water at 1 mg/mL concentration (Figure 7.1d). NDQ20 begins separating within minutes upon standing and forms a large, easily observable pellet and a clear supernatant when centrifuged at 21,130 x g for 30 min.

7.3.2 UVRR Characterization of DQ20

We used UVRR spectroscopy to investigate the secondary structure of DQ20 (Figure 7.1i). We find that the UVRR spectrum (Figure 7.2) and secondary structure (Figure 7.3) of DQ20 is similar to those previously reported for DQ10.[82] DQ20 contains $AmIII_3^S$ peaks at ~1215 cm^{-1}, ~1275 cm^{-1}, and ~1250 cm^{-1}, which result from Ψ angle distributions characteristic of turn-

like (Ψ = 10°), 2.5_1-helix-like (Ψ = 175°), and PPII-like (Ψ = 150°) secondary structures, respectively (Figure 7.3). From this we conclude that DQ20 has a predominantly PPII-like conformation interspersed with turn regions and with terminal residues in a 2.5_1-helix-like conformation, as previously found for DQ10.[82]

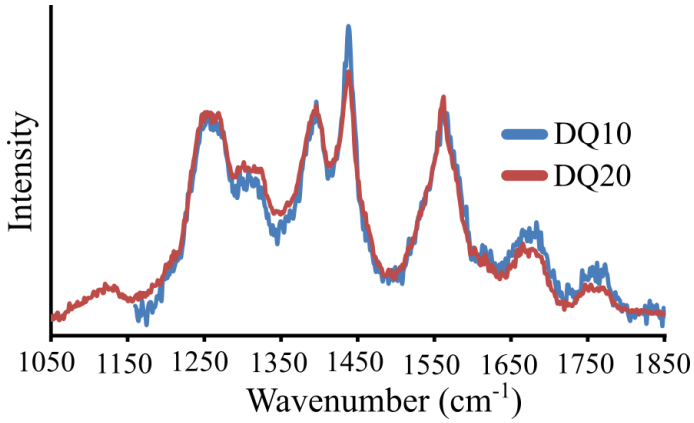

Figure 7.2: UVRR spectra of (blue) DQ10 and (red) DQ20. The DQ10 spectrum was previously measured by Punihaole et al.[82] Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

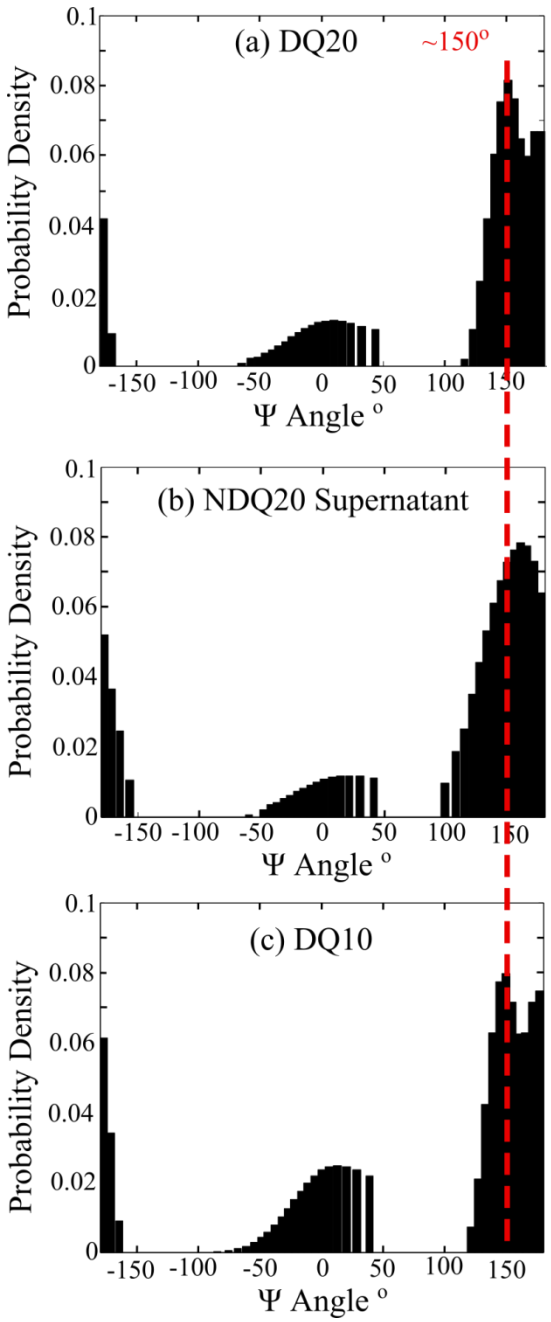

Figure 7.3: UVRR determined Ψ angle distributions of (a) DQ20, (b) NDQ20 supernatant, and (c) DQ10. The Ψ distribution for DQ10 was previously calculated by Punihaole et al.[82] Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

7.3.2.1 DQ20 Contains Gln Side Chains Hydrogen Bonded to Water

DQ20 has overlapping AmI^S and AmI^P bands at ~1677 cm^{-1}. To identify the frequency of the AmI^P band we examined the 197-204 nm difference spectrum of DQ20 (Figure 7.4a). We find that the AmI^P band is located at ~1681 cm^{-1}. Using eq. 7.1 we estimate that the ΔH_{int} is ~-4.1 kcal/mol, which is similar to the ΔH_{int} between the Gln side chain and water.[81] Our results show that DQ20 has Gln side chains that are hydrogen bonded to water (Figure 7.1i).

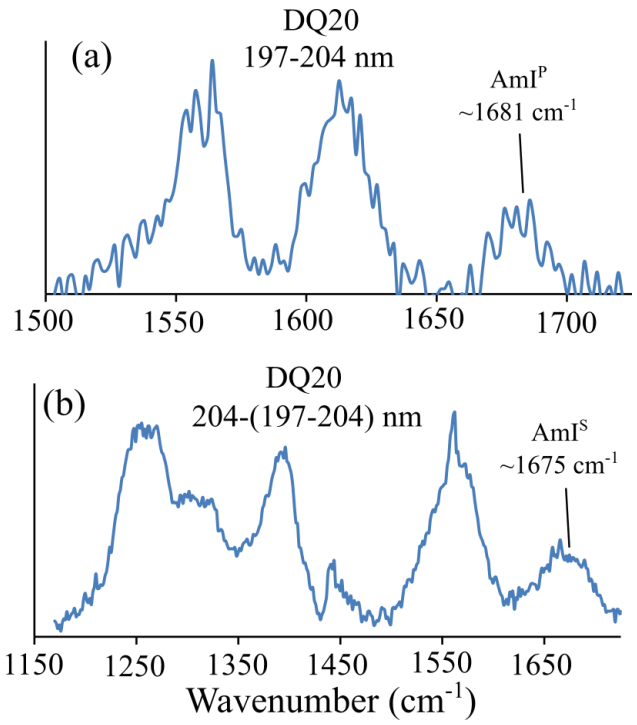

Figure 7.4: (a) 197-204 nm difference spectrum and (b) 204-(197-204) nm difference spectrum of DQ20. The AmI^P band is located at ~1681 cm^{-1} and the AmI^S band is located at ~1675 cm^{-1}. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

We also examined the 204-(197-204) nm difference spectrum of DQ20 to determine its AmI^S frequency (Figure 7.4b). We find that DQ20 contains a broad AmI^S band centered at ~1675 cm^{-1}. Previously, Punihaole et al.[81] observed the frequency of the AmI^S band to be ~1660 cm^{-1} in β-sheet Q10 fibrils with strong backbone-backbone hydrogen bonding. As discussed above, weaker hydrogen bonding upshifts the AmI band frequency. Thus, the fact that the AmI^S frequency of DQ20 is upshifted compared to that of fibrils suggests that the peptide backbone of DQ20 involves weaker hydrogen bonding than that of β-sheet fibrils. This is expected for PPII-like secondary structures that generally have water exposed peptide backbones.

7.3.2.2 Fibrillization of DQ20

At room temperature and low pH (pH = ~+2-3) the DQ20 monomers are stable for >>1 week. However, incubation of DQ20 at 37°C and pH=~+7 results in fibril formation in ~1 week (Figure 7.1j). The UVRR spectra of DQ20 fibrils are similar to those previously observed for DQ10 fibrils (Figure **7**.5a).[46] The $AmIII_3^S$ and AmI bands of DQ10 and DQ20 fibrils are similar, indicating that their secondary structures and hydrogen bonding interactions are similar (Figure C.2). The UVRR spectra of DQ20 fibrils show $AmIII_3^S$ bands at ~1230 cm^{-1} and ~1210 cm^{-1} that derive from Ψ angle distributions centered at ~145° and ~123° (Figure **7**.6a). As discussed previously by Punihaole et al., Ψ angle distributions centered at ~145° are characteristic of antiparallel β-sheet structures, while those centered at ~123° are characteristic of parallel β-sheet structure.

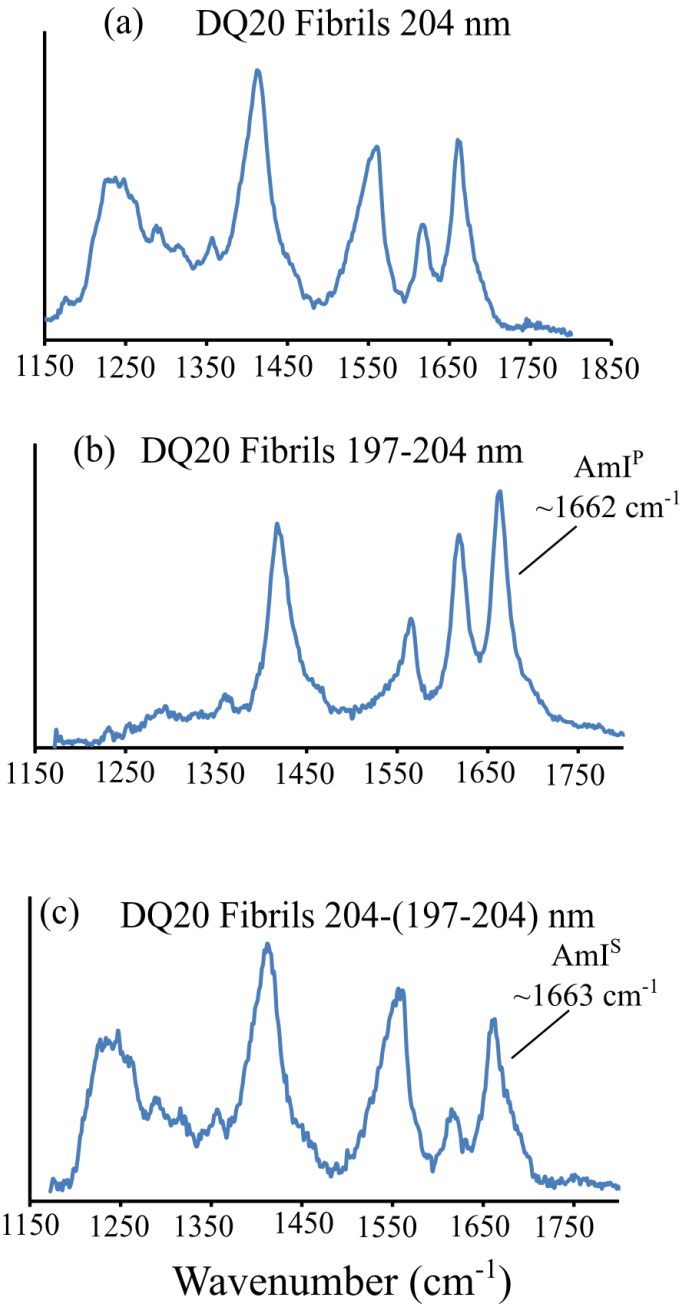

Figure 7.5: (a) 204 nm, (b) 197-204 nm, and (c) 204-(197-204) nm UVRR spectra of DQ20 fibrils. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

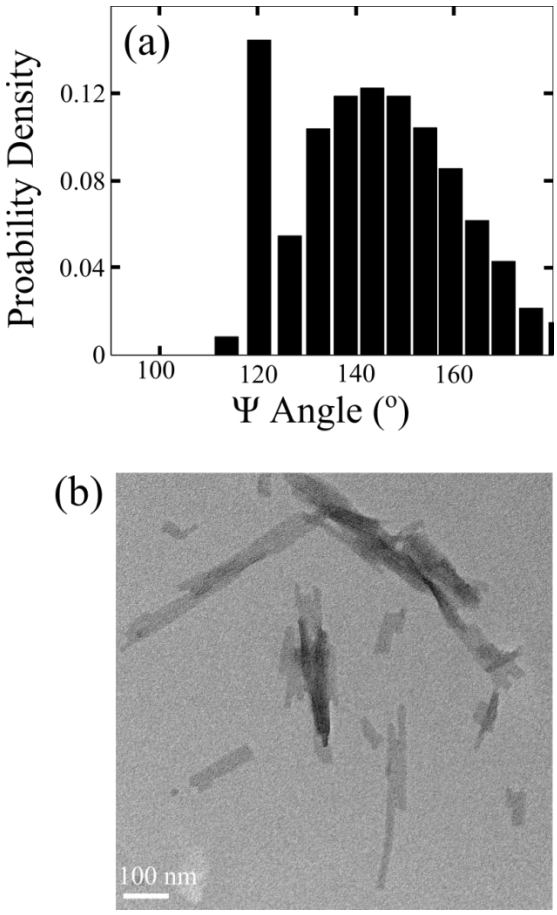

Figure 7.6: (a) Ψ angle distribution and (b) TEM image of DQ20 fibrils. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

In the 197-204 nm difference spectrum, the AmI^P band maximum of DQ20 fibrils occurs at ~1662 cm^{-1} (Figure **7**.5b). As discussed above, this is characteristic of Gln side chain C=O groups involved in strong side chain-side chain and/or side chain-backbone hydrogen bonding. Using eq. 7.1 we estimate that the ΔH_{int} for Gln side chain C=O groups in DQ20 fibrils is ~-5.7 kcal/mol, similar to that previously found for DQ10 fibrils.[81] Similarly, an AmI^S band is observed in the 204-(197-204) nm difference spectrum (Figure **7**.5c) of DQ20 fibrils at ~1663 cm^{-1} that is characteristic of backbone amides involved in β-sheet backbone-backbone hydrogen bonding as previously observed for Q10 fibrils.[46,81] Overall, the AmI^P and AmI^S band frequencies show that both backbone and side chain amide C=O groups in DQ20 fibrils are involved in strong peptide-peptide hydrogen bonding.

7.4 Solubility of NDQ20

As discussed above, NDQ20 does not form clear solutions in water (Figure 7.1d). We used UV absorbance to detect the presence of any soluble peptide in the supernatant after ultracentrifugation of NDQ20. We found that the NDQ20 supernatant contains significant absorbance; suggesting the presence of soluble peptide (Figure 7.7). Using UV absorbance and UVRR spectroscopies we calculate that the NDQ20 supernatant contains ~0.076 mg/ml of peptide. The supernatant peptide concentration calculations are discussed in detail below.

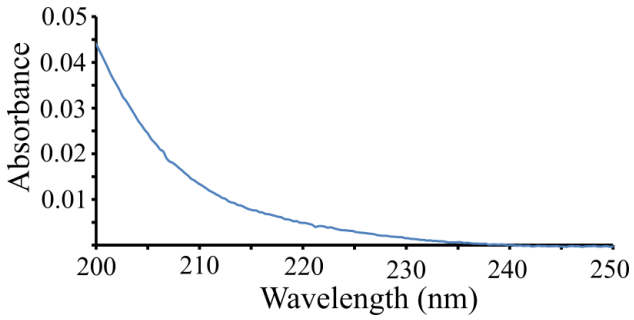

Figure 7.7: Absorption spectrum of NDQ20 supernatant. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

7.4.1 Soluble NDQ20 Supernatant Fraction Contains a PPII-like Peptide Conformation

We used UVRR to examine the structure of the peptide in the NDQ20 supernatant (Figure 7.1e). The UVRR spectrum of the NDQ20 supernatant is similar to that of DQ20 (Figure 7.8). We observe $AmIII_3^S$ bands at ~1250 cm^{-1}, ~1215 cm^{-1}, and ~1275 cm^{-1}, which derive from Ψ angle distributions centered at ~150°, ~10°, and ~175° respectively (Figure 7.3b). As discussed above for DQ20, these Ψ angles are characteristic of PPII-like, turn-like, and 2.5_1-helix-like secondary structures, respectively. From the Ψ angle distributions we conclude that the peptide in the NDQ20 supernatant has the same secondary structure as DQ20. However, the widths of the PPII-like and 2.5_1-helix-like Ψ angle distributions (and $AmIII_3^S$ bands) of the NDQ20 supernatant are larger than that of DQ20 (Figure 7.3). This indicates that the PPII-like and 2.5_1-helix-like structures of the NDQ20 supernatant is less ordered compared to that of DQ10. Also, we find that the AmI^P and AmI^S bands of the NDQ20 supernatant are found at

~1680 cm^{-1}, which is similar to that of DQ20 and indicates Gln side chains that are hydrogen bonded to water.

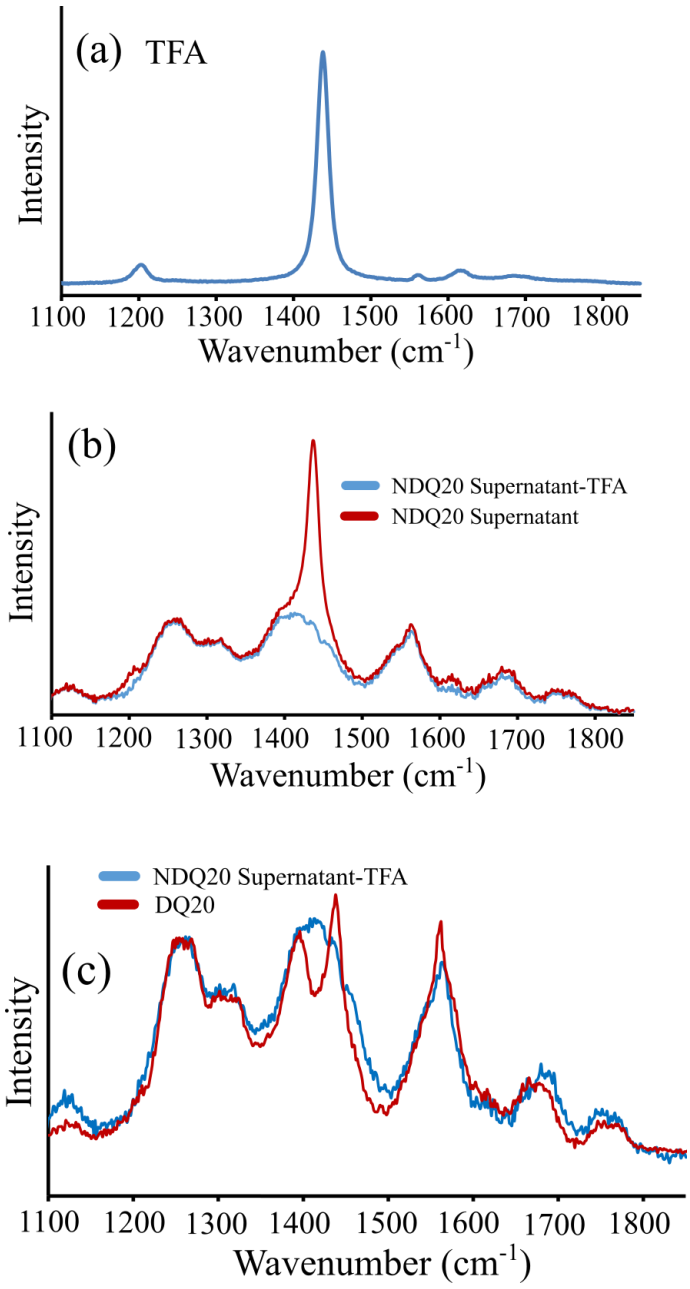

Figure 7.8: UVRR spectra of (a) TFA and (b) the NDQ20 supernatant with (blue) and without (red) subtraction of TFA. (c) UVRR spectral comparison of (red) DQ20 and (blue) the NDQ20 supernatant with TFA subtracted. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763.

7.4.2 TFA Contamination in NDQ20 Supernatant

In the NDQ20 supernatant UVRR spectra, we observe a strong, dominant peak at ~1435 cm^{-1} and a weaker peak at ~1205 cm^{-1} that result from trifluoroacetic acid (Figure 7.8a and b). These bands derive from CO stretching and asymmetric C-F stretching bands of the trifluoroacetate ion.

Thermo Fisher Scientific synthesized our Q20 using an Fmoc solid-phase peptide synthesis where TFA was used for peptide cleavage and as a cosolvent for HPLC purification (see Methods Section). As a result, TFA is a common impurity in our Q20 peptides.[174] When NDQ20 is centrifuged most of it pellets out. However, because of TFA's miscibility with water it remains in the supernatant.

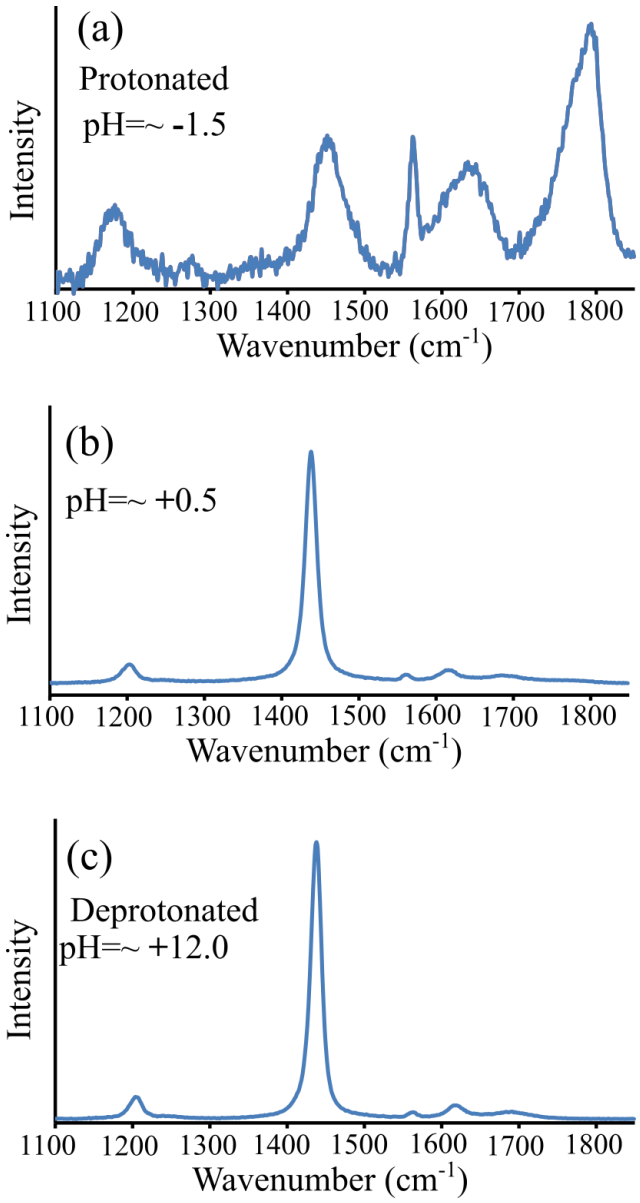

Figure 7.9: UVRR spectra (204 nm) of 10% (v/v) TFA in water at (a) pH=~-1.5, (b) pH=~+0.5, and (c) pH=~+12. At pH=~+12 TFA is deprotonated and at pH=~-1.5 TFA is predominantly protonated. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

Figure 7.9b shows the 204 nm UVRR spectrum of 10% (v/v) TFA (pK_a=~+0.5)[175] in water at pH=~+0.5. This spectrum is consistent with that previously reported for deprotonated TFA (trifluoroacetate).[176,177] Also, the Figure 7.9b UVRR spectrum is identical to that of deprotonated TFA (Figure 7.9c) indicating that the trifluoroacetate ion is selectively resonance enhanced at 204 nm. In contrast, the UVRR spectrum of protonated TFA (Figure 7.9a) significantly differs from that of trifluoroacetate and has a much smaller 204 nm UVRR crosssection.

We assign the spectrum of trifluoroacetate based on the assignments of Robinson and Taylor,[176] and Klemperer and Pimentel.[177] The most intense peak in the TFA UVRR spectrum is located at ~1435 cm^{-1} and is assigned to symmetric carboxylate stretching motion. We also observe peaks at ~1205 cm^{-1} and 1620 cm^{-1} that were previously assigned to asymmetric C-F stretching[176] and asymmetric carboxylate stretching,[176,177] respectively.

We assign our UVRR spectrum of protonated TFA based on the assignments of Fuson et al.[178] For protonated TFA, the most intense UVRR peaks are found at ~1795 cm^{-1} and ~1772 cm^{-1}. We assign these bands to C=O stretching of TFA based on the work by Fuson et al.[178] We also observe a band at ~1452 cm^{-1} that is assigned to C=O deformation.[178] The bands at ~1014 cm^{-1} and 1177 cm^{-1} and ~1273 cm^{-1} are assigned to OH out-of-plane deformation, C-F stretching, and OH in-plane deformations, respectively, of protonated TFA.[178]

We found that the NDQ20 supernatant has a pH of ~+3.5. Therefore TFA is predominantly deprotonated in the NDQ20 supernatant. The peaks at ~1435 cm^{-1} and ~1205 cm^{-1} in the NDQ20 supernatant are assigned to the strongly resonance enhanced carboxylate

stretching and the C-F stretching bands of the trifluoroacetate ion. Subtraction of the TFA UVRR spectrum from that of the NDQ20 supernatant reduces the intensity of the ~1435 cm^{-1} and ~1205 cm^{-1} bands with little effect on the other UVRR bands, including the structurally sensitive $AmIII_3^S$ and AmI bands (Figure 7.8b).

7.4.3 Concentration of TFA in the NDQ20 Supernatant

The NDQ20 supernatant contains both TFA and the Q20 peptide. We determined the concentration of TFA in the NDQ20 supernatant from its UVRR spectrum. To do this we first calculated the Raman cross section of the ~1435 cm^{-1} band of trifluoroacetate using the following equation:

$$\sigma_i = \frac{I_i k_r C_r \sigma_r}{I_r k_i C_i}\left(\frac{A_i + A_{ex}}{A_r + A_{ex}}\right)$$

where σ_i is the Raman cross section of the trifluoroacetate ~1435 cm^{-1} band, σ_r is the Raman cross section of an internal standard Raman band, I_i is the intensity of the ~1435 cm^{-1} trifluoroacetate band, and I_r is the intensity of the internal standard Raman band. The factors k_r and k_i are the spectrometer efficiencies for the Raman bands of trifluoroacetate and the internal standard, respectively, and C_r and C_i are the concentrations of trifluoroacetate and the internal standard, respectively. The term in parentheses approximately corrects for sample self-absorption where A_{ex} is the sample absorbance at the excitation frequency, A_i is the sample absorbance at the trifluoroacetate Raman band of interest, and A_r is the sample absorbance at the internal standard Raman band.

To determine the 204 nm UVRR cross section for the ~1435 cm^{-1} band of trifluoroacetate, we measured the UVRR spectrum of TFA at pH=~+12 using sodium perchlorate ($NaClO_4$) as an internal Raman cross section standard. Raman cross section of $NaClO_4$ was estimated to be ~$1.18x10^{-27}$ cm^2 $molecule^{-1}$ sr^{-1} for 204 nm excitation by extrapolating the Raman cross section measurements of Dudik et al.[179] From eq. 7.2 we calculate that the 204 nm Raman cross section of the ~1435 cm^{-1} trifluoroacetate band is ~$1.08(\pm0.01)x10^{-26}$ cm^2 $molecule^{-1}$ sr^{-1}.

Using the Raman cross section of trifluoroacetate we can determine the concentration of TFA in the NDQ20 supernatant. Rearrangement of eq. 7.2 gives:

$$C_i = \frac{I_i k_r C_r \sigma_r}{I_r k_i \sigma_i}\left(\frac{A_i + A_{ex}}{A_r + A_{ex}}\right) \qquad \textbf{7.3}$$

We collected UVRR spectra of the NDQ20 supernatant using sodium perchlorate as an internal standard. The NDQ20 supernatant has a pH of ~+3.5; thus, essentially all of the TFA in the supernatant will be deprotonated and contribute to the ~1435 cm^{-1} band of trifluoroacetate. Because we know the 204 nm Raman cross section of the ~1435 cm^{-1} trifluoroacetate band, we can calculate the concentration of TFA in the NDQ20 supernatant using eq. 7.3. We find that the concentration of TFA in the NDQ20 supernatant is ~810±70 μM.

7.4.4 Concentration of Peptide in the NDQ20 Supernatant

Using the concentration of TFA in the supernatant, calculated above, we determined the absorbance due to TFA at 214 nm in the NDQ20 supernatant. We constructed a calibration curve

for TFA in the NDQ20 supernatant at 214 nm using a standard addition method (Figure 7.10a). From the calibration curve we calculate that 810 μM TFA should result in an absorbance of ~0.0006 at 214 nm. For the NDQ10 supernatant we find an absorbance of ~0.0113±0.002 at 214 nm (Figure 7.7). Thus, the absorbance of the peptide in the NDQ20 supernatant is ~0.0107. We constructed a calibration curve using highly soluble DQ20 peptide to determine the concentration of peptide in the supernatant (Figure 7.10b). At 214 nm, DQ20 has a molar absorptivity of ~19220 M^{-1} cm^{-1}. From the DQ20 calibration curve we calculate that the NDQ20 supernatant contains ~0.076 mg/mL (~24.8 μM) of the Q20 peptide.

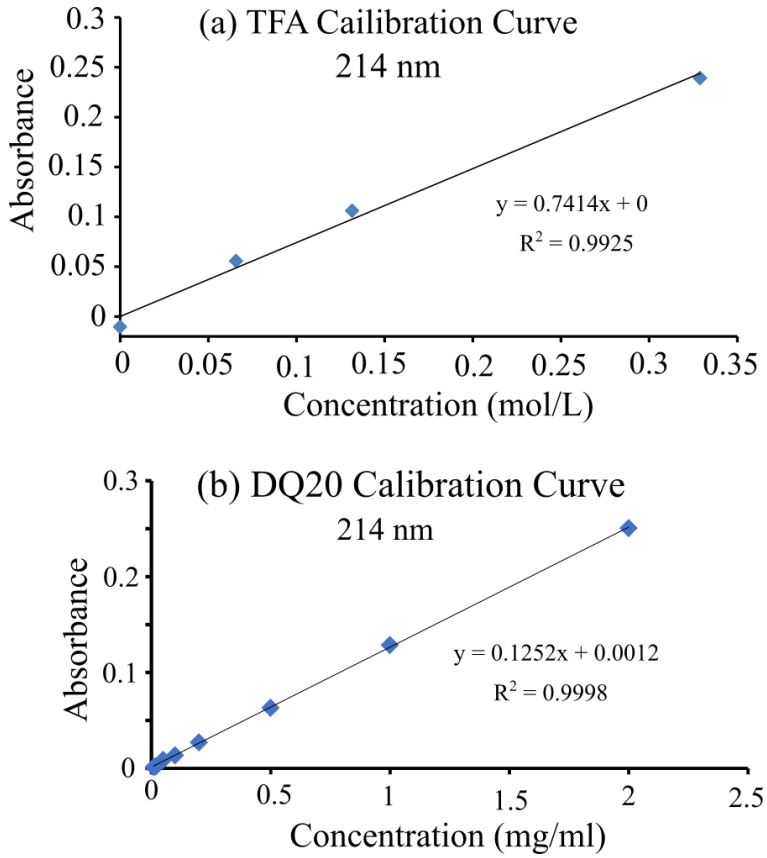

Figure 7.10: Absorbance calibration curve at 214 nm for (a) TFA and (b) DQ20. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

7.4.5 TFA used in Solid-Phase Peptide Synthesis may Induce PPII-Like Conformation

Our results show that the peptide in the NDQ20 supernatant has a PPII-like structure similar to that of disaggregated peptides and contains TFA. This is consistent with, Thakur and coworkers[180,181] who showed that TFA alone causes disaggregation of polyQ peptides. Further, disaggregated polyQ peptides have previously been found to occur in PPII-like conformations.[82] Thus, we hypothesize that the PPII-like structure observed in the NDQ20 supernatant results from the use of TFA during peptide synthesis. However, the mechanism by which TFA induces structural change in polyQ peptides remains poorly understood.

7.5 Structure of SPPS Q20 and NDQ20

7.5.1 UVRR Characterization and Ψ Angle Distribution of SPPS Q20

The UVRR spectrum of SPPS Q20 is similar to that of NDQ10[82], which was found to have a collapsed β-strand-like conformation (Figure 7.11).[82] The SPPS Q20 UVRR spectrum shows an $AmIII_3^S$ frequency of ~1245 cm^{-1}, which is 5 cm^{-1} upshifted compared to that of NDQ10 (~1240 cm^{-1}). This ~5 cm^{-1} $AmIII_3^S$ frequency difference between SPPS Q20 and NDQ10 likely results from differences in peptide backbone hydration.

Mikhonin et al. previously correlated the $AmIII_3^S$ frequency to the Ψ angle for peptides in a variety of backbone solvation states.[64] They found that the frequency of the $AmIII_3^S$ band is sensitive to water solvation of the peptide backbone. Comparison of the $AmIII_3^S$ frequency–Ψ angle correlations for crystalline peptides and solvated peptides shows an ~4 cm^{-1} upshift of the $AmIII_3^S$ band in peptide crystals compared to peptides in aqueous solution. This is similar to the ~5 cm^{-1} upshift of the $AmIII_3^S$ band we observe for SPPS Q20 compared to aqueous NDQ10. Thus, the $AmIII_3^S$ frequency differences between NDQ10 and SPPS Q20 are most likely to result from water exposure of the peptide backbone. In agreement, the AmI^P band of SPPS Q20 indicates that the side chain amide groups are not hydrogen bonded to water (see below) while NDQ10 side chain and backbone carbonyl groups are hydrogen bonded to water.[82] See below for a detailed discussion of the SPPS Q20 AmI bands.

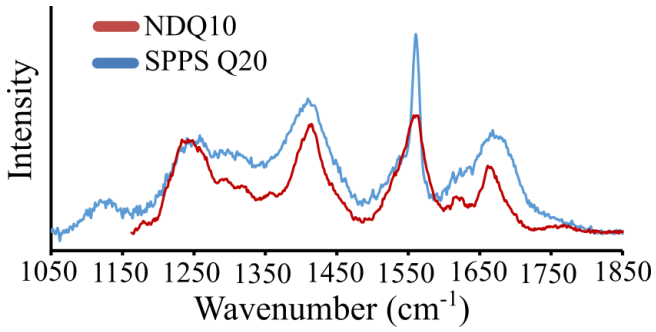

Figure 7.11: UVRR spectra of (red) NDQ10 and (blue) SPPS Q20. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

Punihaole et al.[82] previously calculated the Ψ angle distribution of the NDQ10 peptide from the $AmIII_3^S$ band using an equation derived by Mikhonin et al.[64] for peptides in aqueous

solution with an unknown hydrogen bonding environment (eq. C.4). Using this equation, Punihaole et al. calculated that aqueous NDQ10 has a Ψ angle distribution centered at ~140°, characteristic of a β-strand-like secondary structures.[82]

Here, we use the $AmIII_3^S$ band of SPPS Q20 to calculate its Ψ angle distribution. However, because the $AmIII_3^S$ frequency depends on both the Ψ angle and hydrogen bonding to the amide C=O and N-H groups,[64,90] we must account for the $AmIII_3^S$ frequency shift resulting from differences in hydrogen bonding between dissolved NDQ10 and SPPS Q20. Therefore, instead of using C.4 for aqueous peptides, we used the measured $AmIII_3^S$ frequency-Ψ angle correlation for crystalline peptides derived by Mikhonin et al.[64] (equation C.5). This equation best models the hydrogen bonding and solvation environments expected for SPPS Q20. Using this equation we calculate that SPPS Q20 has a Ψ angle distribution centered at ~138°, which is characteristic of β-strand-like structures (Figure 7.12b) such as that observed for NDQ10 (Ψ = ~140°).[82] Also, it is important to note that the $AmIII_3^S$ band and Ψ angle distribution of SPPS Q20 is significantly broader than that of NDQ10. This indicates that the β-strand-like conformation of SPPS Q20 is less ordered and contains more conformational variations compared to that of NDQ10.

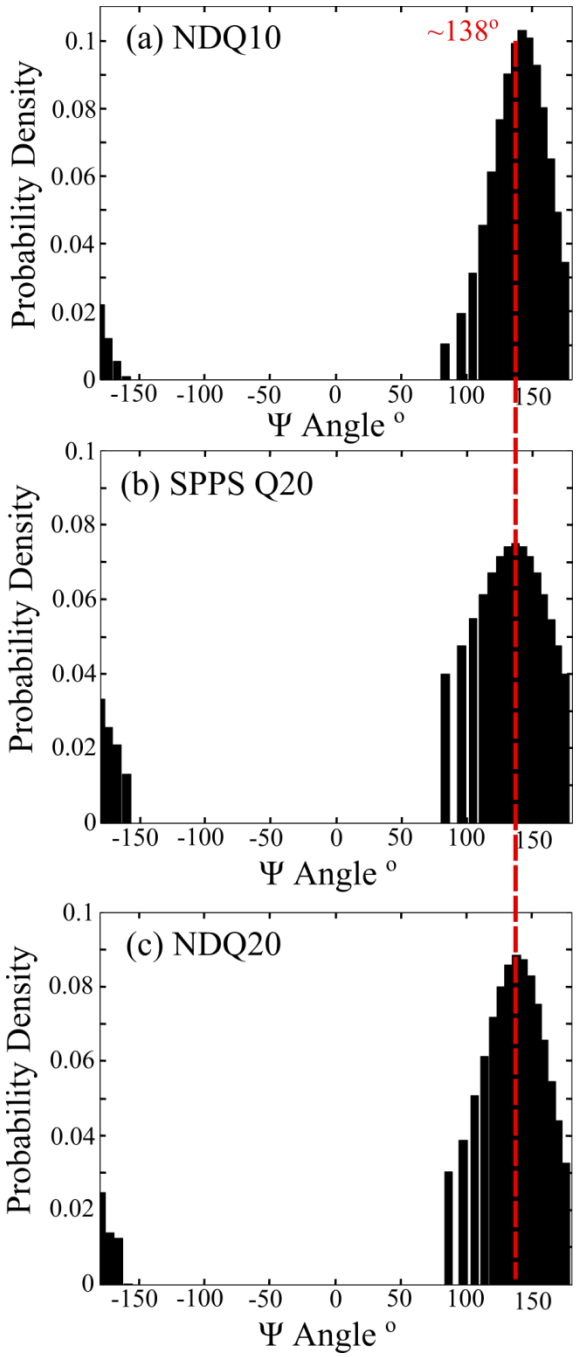

Figure 7.12: Ψ angle distribution of (a) NDQ10 (b) SPPS Q20, and (c) NDQ20. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

7.5.2 UVRR Characterization and Ψ Angle Distribution of NDQ20

We investigated the structure of NDQ20 (Figure 7.1d) by measuring its UVRR spectrum (Figure 7.13). We find that the UVRR spectrum of NDQ20 in water is very similar to that of SPPS Q20 (Figure 7.12b) with an $AmIII_3^S$ band at ~1246 cm^{-1}. Using eq. C.5, we calculate that the NDQ20 Ψ angle distribution is centered at ~138°, which is characteristic of a β-strand-like conformation (Figure 7.12c).

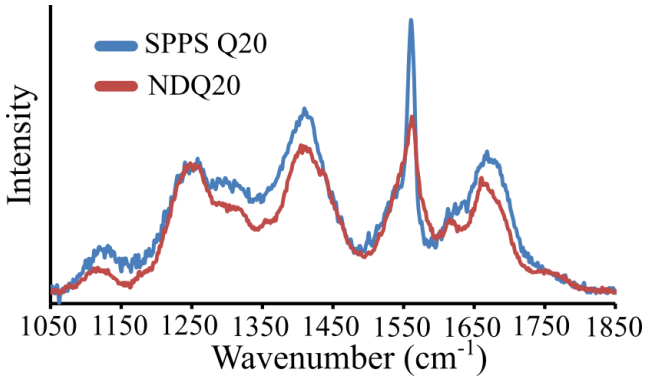

Figure 7.13: UVRR spectra of (blue) SPPS Q20 and (red) NDQ20 in water. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

7.5.3 Side Chain and Backbone Hydrogen Bonding in SPPS Q20 and NDQ20

At 204 nm excitation, SPPS Q20 and NDQ20 both contain broad (~70 cm^{-1} FWHH) AmI bands centered at ~1670 cm^{-1} (Figure 7.14). In contrast to SPPS Q20, NDQ20 also contains a narrow AmI peak at ~1660 cm^{-1}, which indicates strong peptide-peptide hydrogen bonding.

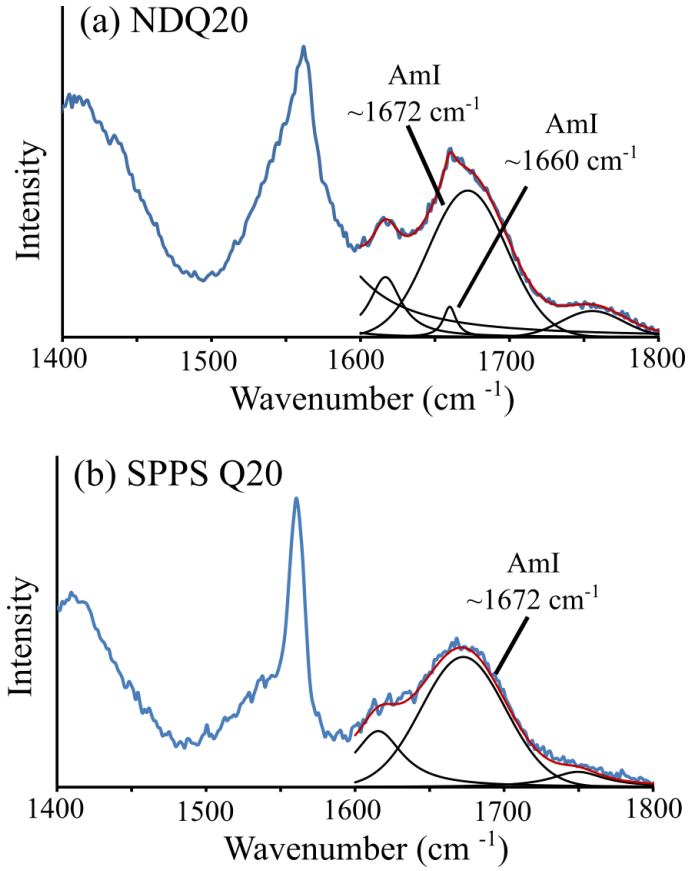

Figure 7.14: UVRR AmI spectral region of (a) NDQ20 and (b) SPPS Q20. The spectra were modeled as a sum of Gaussian and Lorentzian bands, shown in black, as described in Appendix C. The AmI bands are labeled in each spectrum. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

We examined the 197-204 nm difference spectra of SPPS Q20 and NDQ20 to highlight the contributions of the AmI^P bands (Figure 7.15a and c). We find that SPPS Q20 and NDQ20 both contain narrow AmI^P bands located at ~1666 cm^{-1} and ~1662 cm^{-1} respectively. These frequencies are indicative of strong side chain-side chain and/or side chain-backbone hydrogen bonding such as that found in polyQ fibrils.[46] Using eq. 7.1 we estimate that the Gln side chain ΔH_{int} = ~-5.3 kcal/mol for SPPS Q20 and ~-5.6 kcal/mol for NDQ20.

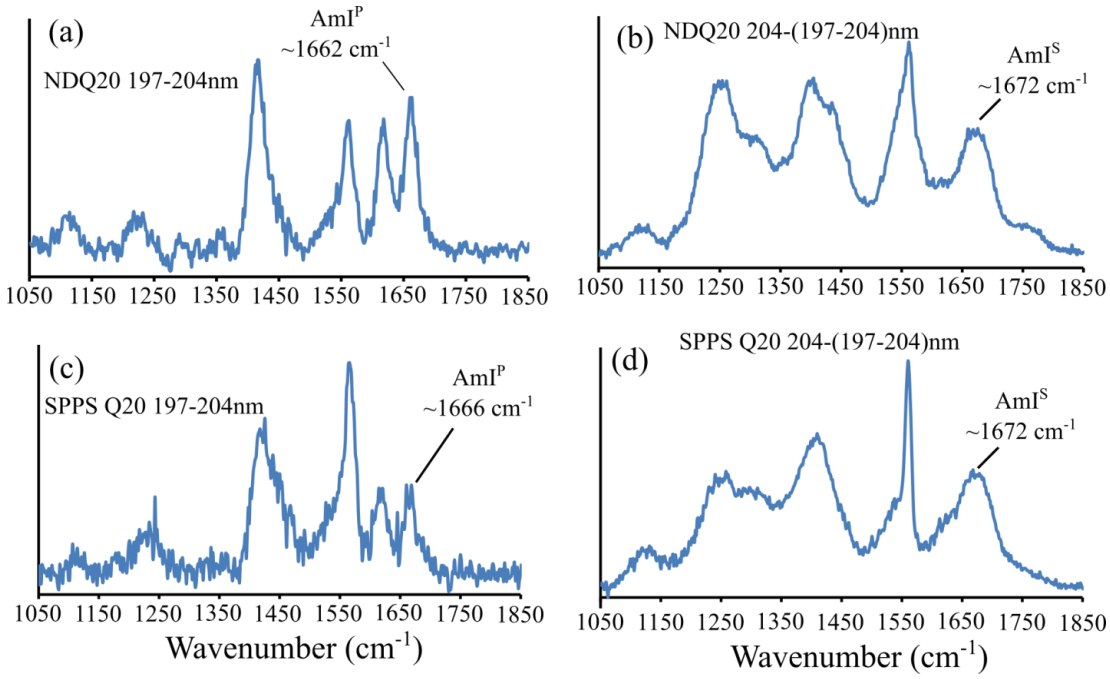

Figure 7.15: 197-204nm UVRR difference spectra of (a) NDQ20 and (c) SPPS Q20, and 204-(197-204)nm difference spectra of (b) NDQ20 and (d) SPPS Q20. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

The increased relative intensity of the ~1662 cm^{-1} AmI^P band in the UVRR spectra of NDQ20, compared to that of SPPS Q20, indicates that NDQ20 contains a larger population of Gln side chains involved in strong side chain-peptide hydrogen bonding. Also, the lack of an ~1680 cm^{-1} AmI^P band in NDQ20 and SPPS Q20 indicates that neither species contains a significant population of Gln side chains hydrogen bonded to water.

From the 204-(197-204) nm spectra (Figure 7.15b and d), we find that both NDQ20 and SPPS Q20 have an AmI^S frequency of ~1675 cm^{-1}. This frequency is upshifted from that of polyQ fibrils (~1660 cm^{-1}) indicating that NDQ20 and SPPS Q20 contain weaker backbone-peptide hydrogen bonding. Also, the AmI^S bands of NDQ20 and SPPS Q20 are broad (~70 cm^{-1} FWHH) suggesting that the peptide backbone amides are found in a variety of hydrogen bonding environments that inhomogenously broaden the AmI^S band.

7.6 Stability and Activation Barrier Between the PPII-Like and β-Strand-Like Structures of PolyQ

Punihaole et al.[82] previously showed that the PPII-like and β-strand like conformations do not interconvert because of a high activation barrier between them. We find that the PPII-like polyQ structure of DQ10[82] and DQ20 contains negligible peptide-peptide hydrogen bonding and extensive peptide-water hydrogen bonding. PPII conformations are thought to be stabilized by peptide backbone-water hydrogen bonding[182,183] and/or the low-energy water-water hydrogen bonding structure of the PPII peptide solvation shell. [29,182,184–188] Thus, it is likely that the polyQ PPII-like structure is also stabilized by these interactions. A PPII-like → β-strand-like polyQ

structural transition would have to overcome the energy barrier(s) associated with breaking peptide backbone-water hydrogen bonds and/or disrupting the water-water hydrogen bonding of the PPII solvation shell.

In contrast, the β-strand-like structure of NDQ10[82] and NDQ20 contains significant side chain-peptide hydrogen bonding. β-strand structures are stabilized by peptide-peptide hydrogen bonding between neighboring β-strands. [21] Thus, polyQ β-strand conformations are likely stabilized by their side chain-peptide hydrogen bonding. A β-strand-like → PPII-like polyQ structural transition would have to overcome the energy barrier associated with breaking peptide-peptide hydrogen bonds.

7.7 NDQ20 Fibrillization

To determine if NDQ20 (Figure 7.1d) will from fibrils, we examined the UVRR spectrum of aqueous NDQ20 after incubation for ~5 hours at room temperature (~18°C) and low pH (pH = ~+2-3) (Figure 7.1f). Incubation of NDQ20 resulted in a UVRR spectrum similar to that of NDQ10 fibrils reported by Punihaole et al. (Figure 7.16b).[46] After incubation, the $AmIII_3^S$ band downshifts to ~1233 cm^{-1}, which is characteristic of the antiparallel-β-sheet structure of polyQ fibrils (Ψ = ~148°) (Figure 7.17b) that contain dry fibril cores.[46] We also observe a low intensity $AmIII_3^S$ band at ~1210 cm^{-1} that corresponds to a Ψ angle distribution centered at ~123°. As discussed previously by Punihaole et al., Ψ angles of ~123° arise from sub-populations of parallel β-sheet in polyQ fibrils.[46] The widths of both the parallel and antiparallel β-sheet Ψ angle distributions are larger for NDQ20 fibrils compared to NDQ10 fibrils. This

indicates that the secondary structure of the NDQ20 fibrils is less ordered compared to NDQ10 fibrils.

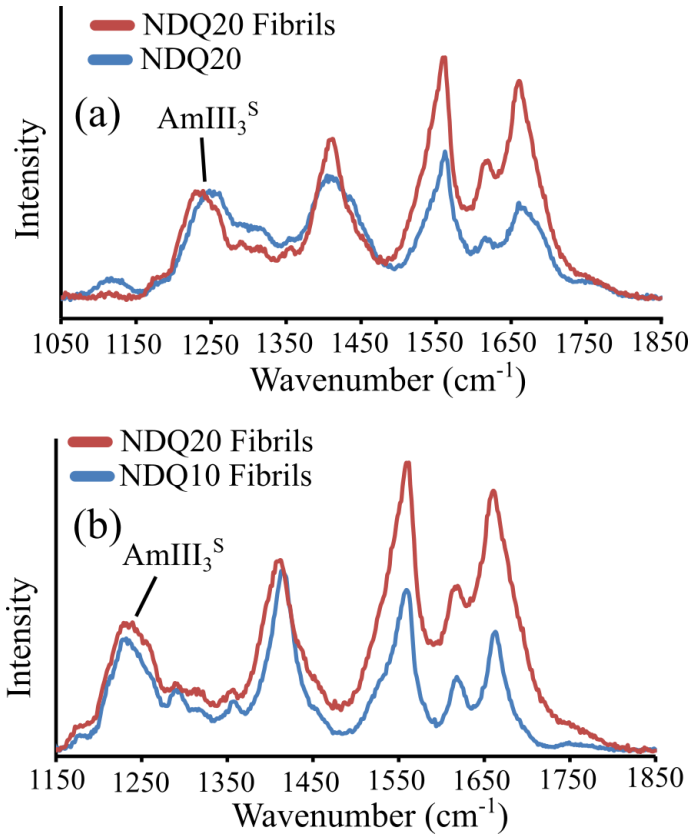

Figure 7.16: (a) UVRR spectra of NDQ20 (blue) and NDQ20 fibrils (red). (b) Comparison of the UVRR spectra of NDQ20 fibrils (red) and NDQ10 fibrils (blue) previously collected by Punihaole et al.[46] Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

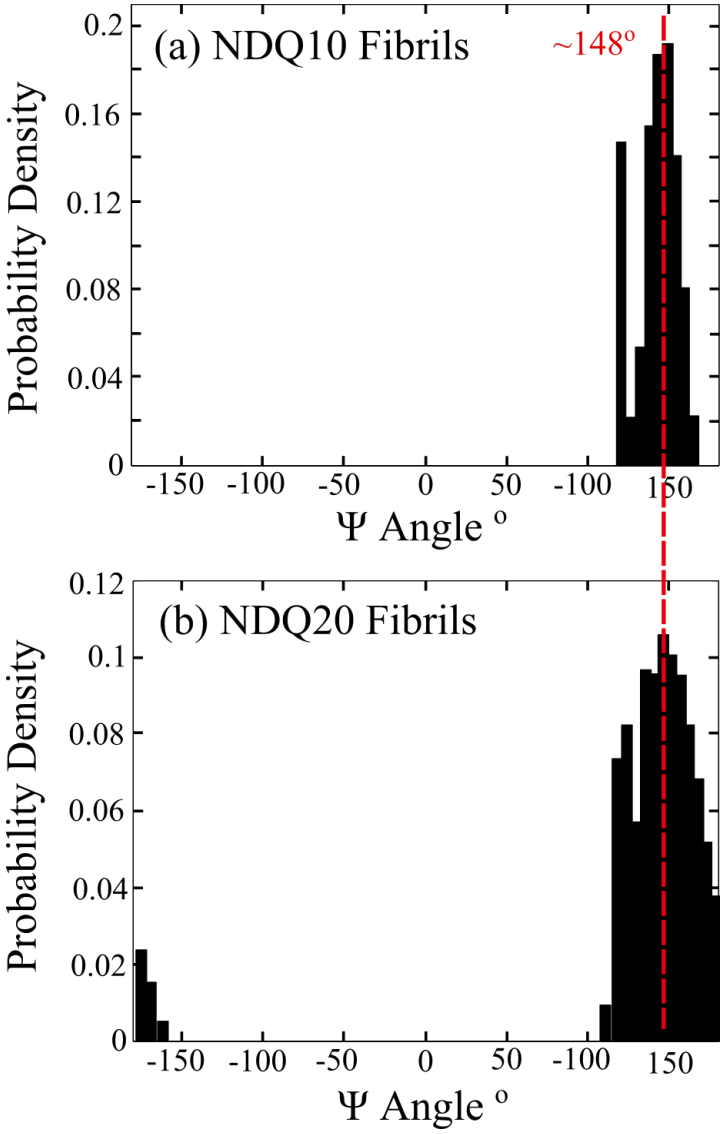

Figure 7.17: Ψ angle distribution of (a) NDQ10 fibrils and (b) NDQ20 fibrils. Both NDQ10 fibrils and NDQ20 fibrils contain Ψ angle distributions that peak at ~148° and ~123°, which are characteristic of antiparallel β-sheet and parallel β-sheet conformations respectively. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

Additionally, the AmI band downshifts to ~1660 cm^{-1} and increases in relative intensity as a result of formation of strong peptide-peptide hydrogen bonds, as previously observed in Q10 fibril formation.[46] To determine the frequency of the AmI^P and AmI^S bands, we examined the 197-204nm and 204-(197-204) nm UVRR spectra of NDQ20 after incubation (Figure 7.18a and b). We find that the AmI^P band is located at ~1665 cm^{-1}, which corresponds to a side chain ΔH_{int} of ~-5.4 kcal/mol. This is similar to that previously observed for Q10 fibrils.[46,81] Also, we find that the AmI^S band is located at ~1665 cm^{-1}, which suggests strong backbone-backbone hydrogen bonding, as previously observed in polyQ fibrils.[46]

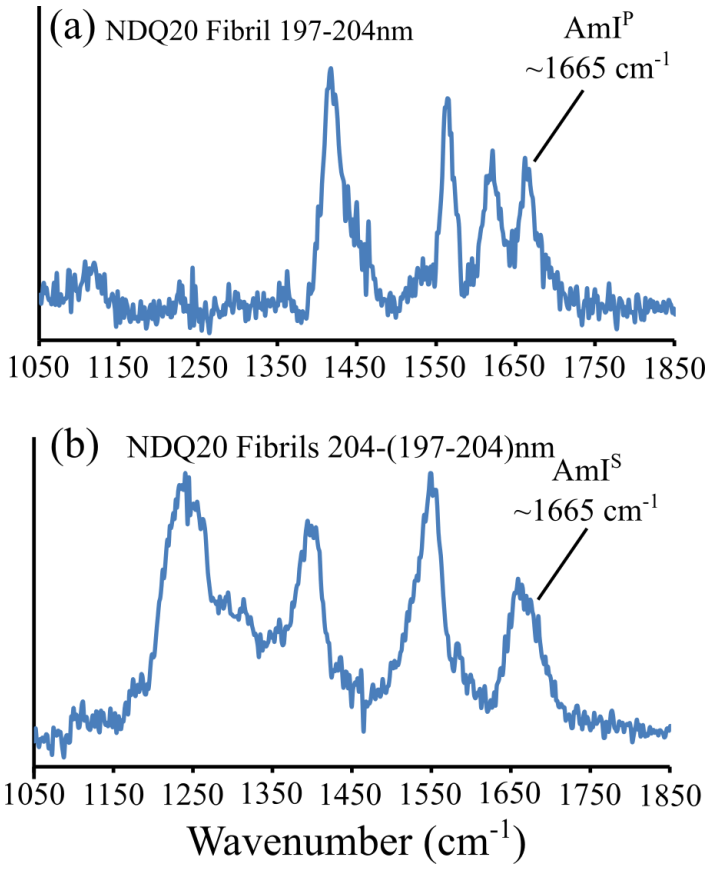

Figure 7.18: (a) 197-204 nm and (b) 204-(197-204) nm difference spectra of NDQ20 after ~5 hours of incubation. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

To confirm formation of fibrils, we collected TEM images of NDQ20 after incubation for ~1-2 days at room temperature and low pH (pH = ~+2-3). Figure 7.19 shows TEM images of NDQ20 fibrils and NDQ10 fibrils. The TEM images of NDQ10 fibrils were previously collected by Punihaole et al.[46] We find that NDQ20 contains amyloid-like fibril aggregates with a similar morphology to that observed for NDQ10.[46] These images confirm that NDQ20 forms fibrils when incubated at room temperature (~18°C) and low pH (pH = ~+2-3).

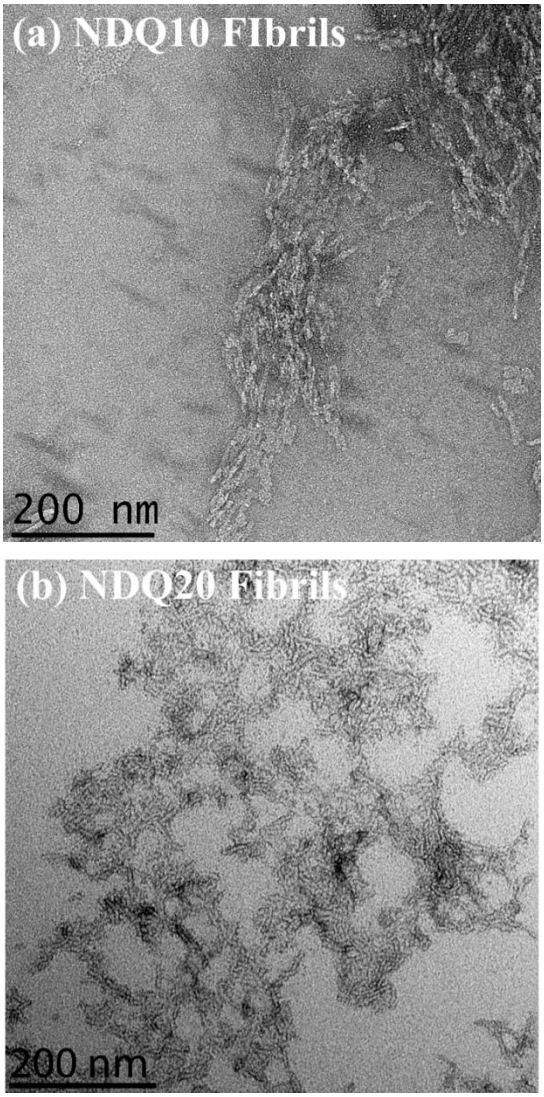

Figure 7.19: TEM images of (a) NDQ10 fibrils and (b) NDQ20 fibrils (Figure 7.1f). Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

7.8 Comparison to Other Results

Our insights into the structures of SPPS Q20 and NDQ20 generally agree with results recently published by Burra and Thakur.[189] They investigated the structures of the insoluble polyQ peptide $K_2Q_9PGQ_4AQ_4PGQ_9PGQ_9K_2$ (PGQ_9A) in the solid-phase synthesis lyophilized powder (SPPS PGQ_9A) and non-disaggregated (NDPGQ_9A) forms using FTIR and CD spectroscopies.

Burra and Thakur[189] found that SPPS PGQ_9A contains predominantly antiparallel β-sheet structure with minority populations of random coil and turn conformations. They conclude that SPPS PGQ_9A must have weaker peptide-peptide hydrogen bonding compared to β-sheet fibrils because turn and random coil conformations have a less optimal hydrogen bonding pattern.[189] This is in agreement with our UVRR data showing that SSPS Q20 is in a β-strand-like conformation with weaker hydrogen bonding compared to fibrils.

In contrast to SPPS PGQ_9A, Burra and Thakur showed that NDPGQ_9A contains only antiparallel β-sheet conformation.[189] Since random coil and turn conformations were not observed in NDPGQ_9A, they concluded that NDPGQ_9A must contain stronger peptide-peptide hydrogen bonding compared to SPPS PGQ_9A.[189] Their conclusions agree with our results showing that NDQ20 contains a β-strand-like structure and stronger peptide-peptide hydrogen bonding compared to SPPS Q20.

Additionally, Burra and Thakur do not observe signs of NDPGQ_9A fibrillization after incubation in water for 3 hours.[189] In contrast, we find that NDQ20 forms fibrils after ~5 hours of incubation in water. It is possible that Burra and Thakur did not observe fibrillization of NDPGQ_9A because it requires an incubation time longer than 3 hours to form fibrils.

The work by Burra and Thakur[189] qualitatively examined the secondary structures of SPPS PGQ_9A and $NDPGQ_9A$. From the secondary structures, they predicted the hydrogen bonding in each species. Our measurements quantitatively extend the work of Burra and Thakur[189] by quantifying the peptide Ψ angle distribution and by directly measuring the peptide backbone and Gln side chain hydrogen bonding interactions for SPPS Q20 and NDQ20. As described above, our measurements provide new insights into the specific secondary structures and hydrogen bonding interactions experienced by insoluble polyQ species.

7.9 Implication of NDQ20 Fibrillization on the PolyQ Fibrillization Mechanism

Currently, there are two major models for the polyQ fibrillization mechanism. The first model, proposed by Wetzel and coworkers[47,58], argues that polyQ aggregation occurs via nucleated growth. In this mechanism, the nucleus is thought to be a thermodynamically unfavorable conformation of the peptide monomer, and fibrils elongate by the recruitment of monomeric units to the growing fibril.[36,47]

The second model for the polyQ aggregation mechanism, developed by Pappu and coworkers, proposes that polyQ peptides form inter- and intra-peptide hydrogen bonds that lead to the formation of non-fibrillar polyQ aggregates.[37,38,43] They propose that these aggregates can undergo a conformational change to form fibrils.[61]

Here, we show that the insoluble NDQ20 (Figure 7.1d) peptide will convert from non-fibril aggregates to fibrils in water. This result suggests that non-fibril polyQ aggregates in water can undergo a conformational transition into fibrils.

7.10 Conclusion

The results of this work are summarized in Table 7.4.

Table 7.4: Summary of Secondary Structure and Hydrogen Bonding for Forms of Q20 Examined in this Study. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

	AmI^P υ (cm^{-1})	Side Chain H-Bonding	AmI^S υ (cm^{-1})	Backbone H-Bonding	Ψ Angle (°)	Secondary structure
DQ20	~1681	S.C.-W.	~1675	B.B.-W.	~10, ~175, ~150	Turn, 2.5_1-Helix, PPII
DQ20 Fibrils	~1662	S.C.-P.	~1663	B.B.-P.	~145, ~123	P-β-Sheet, A-β-Sheet
NDQ20	~1662	S.C.-P.	~1672	B.B.-P.	~138	β-Strand
NDQ20 Supernatant	~1680	S.C.-W.	~1680	B.B.-W.	~10, ~175, ~150	Turn, 2.5_1-Helix, PPII
SPPS Q20	~1666	S.C.-P.	~1672	B.B.-P.	~138	β-Strand
NDQ20 fibrils	~1665	S.C.-P.	~1665	B.B.-P.	~148, ~123	P-β-Sheet, A-β-Sheet

S.C., side chain; W., Water; P., Peptide; B.B., Backbone; P-β-Sheet, parallel β-Sheet; A-β-Sheet, antiparallel β-Sheet

We used UVRR spectroscopy to investigate the structures of Q20. NDQ20 is essentially insoluble in water (Figure 7.1d). In contrast, we find that disaggregation of Q20 renders the peptide (DQ20) highly soluble (Figure 7.1i). Using UVRR spectroscopy we find that DQ20 is in a PPII-like conformation with backbone and side chain amide groups hydrogen bonded to water.

To examine the solubility of NDQ20 (Figure 7.1d), we collected UV absorption spectra of the NDQ20 supernatant (Figure 7.1e) after ultracentrifugation. We find that the supernatant has weak absorbance. Using UV absorbance and UVRR we find a concentration of ~0.076 mg/mL for the peptide in the NDQ20 supernatant. Using UVRR spectroscopy we find that the

soluble fraction of NDQ20 is in a PPII-like conformation (Figure 7.1e) with Gln side chains and peptide backbone hydrogen bonded to water, similar to that of DQ10[82] and DQ20. The presence of PPII-like peptide structure in the NDQ20 supernatant could result from the use of TFA in the peptide synthesis, which can disaggregate polyQ in the absence of HFIP.[180,181]

We used UVRR spectroscopy to investigate the structures of SPPS Q20 (Figure 7.1c) and NDQ20 (Figure 7.1d). We find that SPPS Q20 and NDQ20 are in β-strand-like conformations (Ψ=~138°) similar to that previously observed for NDQ10.[82] Using UVRR spectroscopy, we also probed the backbone and side chain amide hydrogen bonding interactions in SPPS Q20 and NDQ20. We find that both NDQ20 and SPPS Q20 contain backbone amides with weaker inter- and intra-peptide hydrogen bonding compared to that found in fibrils. In contrast the side chain amide groups in SPPS Q20 and NDQ20 are involved in strong side chain-peptide hydrogen bonding. The number of side chain-peptide hydrogen bonds in NDQ20 is greater than that of SPPS Q20.

Upon incubation for ~5 hrs, we find that NDQ20 converts to fibrils (Figure 7.1f).[46] These fibrils have a β-sheet secondary structure with side chain and backbone amides involved in strong peptide-peptide hydrogen bonding. This result, along with the observation that NDQ20 contains more side chain-peptide hydrogen bonds compared to SPPS Q20, suggests that high molecular weight, non-fibrillar polyQ aggregates can undergo a conformational transition into fibrils. The discovery that apparently insoluble polyQ peptides can form fibrils may allow for fibrillization studies on NDQ peptides in a β-strand-like conformation without the need for disaggregation.

7.11 Acknowledgements

Funding for this work was provided by the University of Pittsburgh and the Defense Threat Reduction Agency (DTRA) HDTRA1-09-14-FRCWMD (RSJ, SEW, SAA). SEW gratefully acknowledges support through the Molecular Biophysics and Structural Biology NIH Training Grant (T32 GM 088119).

7.12 Supporting Information

Supporting information can be found in Appendix C and includes the following: UVRR spectral fitting, Ψ angle distribution calculations, TFA absorption spectra, and a comparison between the UVRR spectra of DQ10 and DQ20 fibrils.

8.0 Polyglutamine Solution-State Structural Propensity is Repeat Length Dependent

Adapted with permission from: Ryan S. Jakubek, Riley J. Workman, Stephen E. White, and Sanford A. Asher. Polyglutamine Solution-State Structural Propensity is Repeat Length Dependent. *The Journal of Physical Chemistry B*, 2019, 123(19), 4193-4203.

Author Contributions: R.S.J. collected, analyzed, and interpreted the Raman and CD data with the assistance of S.E.W. R.J.W. performed the computational work and interpreted the data with assistance by R.S.J. and S.E.W. The manuscript was prepared by R.S.J. and S.A.A. with assistance by R.J.W. and S.E.W.

Expanded polyglutamine (polyQ) tracts in proteins, which are known to induce their aggregation, are associated with numerous neurodegenerative diseases. Longer polyQ tracts correlate with faster protein aggregation kinetics and a decreased age of onset for polyQ disease symptoms. Here, we use UV resonance Raman spectroscopy, circular dichroism spectroscopy, and metadynamics simulations to investigate the solution-state structures of the $D_2Q_{15}K_2$ (Q15) and $D_2Q_{20}K_2$ (Q20) peptides. Using metadynamics, we explore the conformational energy landscapes of Q15 and Q20 and investigate the relative energies and activation barriers between these low-energy structures. We compare the solution-state structures of $D_2Q_{10}K_2$ (Q10), Q15, and Q20 to determine the dependence of polyQ structure on the Q tract length. We show that these peptides can adopt two distinct monomeric conformations: an aggregation-resistant PPII-like conformation and an aggregation-prone β-strand-like conformation. We find that longer polyQ peptides have an increased preference for the aggregation-prone β-strand-like conformation. This preference may play an important role in the increased aggregation rate of

longer polyQ peptides that is thought to lead to decreased neurodegenerative disease age of onset for polyQ disease patients.

8.1 Introduction

The expansion of CAG codon repeats in DNA encodes for elongated polyglutamine (polyQ) tracts in proteins.[1] Expanded polyQ tracts in proteins and peptides induce aggregation and fibrillization. This aggregation is associated with numerous neurodegenerative diseases, including Huntington's disease.[1] However, the identity of the toxic species is still debated.[14,16,190,191]

For polyQ diseases, the length of the expanded polyQ tract affects the severity of the disease. Patients with longer polyQ tracts have an earlier disease age of onset, and disease symptoms are only evident if the protein's polyQ tract surpasses a critical length.[5,47,167] For example, clinical presentation of Huntington's disease is only observed in patients with a polyQ tract ≥36 residues long in the huntingtin protein.[1] Additionally, *in vitro* studies of polyQ peptides show that longer polyQ tracts have increased aggregation rates that may be related to the disease age of onset.[47,192]

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 length. Most experimental[34–36,41] and computational[37–39,99] studies conclude that solution-state polyQ peptides are intrinsically disordered regardless of the polyQ tract length. However, some studies suggest the existence of small populations of secondary structure that may play a role in aggregation and cytotoxicity.[6,10,11]

UV resonance Raman (UVRR) spectroscopy has provided detailed insight into the secondary structures of polyQ peptides.[46,82,193] UVRR spectroscopy is a powerful tool for studying protein structure, solvation, and hydrogen bonding interactions.[62,63,81] Excitation in the deep UV (~200 nm) selectively enhances vibrations of the secondary amide peptide backbone and primary amide glutamine (Gln) side chains.[57,194] These resonance-enhanced bands are sensitive to the structure and environment of the peptide backbone and Gln side chains.[63]

Recently, Punihaole et al. used UVRR spectroscopy and metadynamics simulations to investigate the solution-state structures of $D_2Q_{10}K_2$ (Q10).[82] They found that when Q10 is dissolved in water it exists in a well-defined, collapsed β-strand-like conformation. This peptide form is referred to as non-disaggregated Q10 (NDQ10). In contrast, when disaggregated using the methodologies of Chen et al.[31], Q10 exists in a predominantly polyproline II (PPII)-like structure. This peptide conformation is referred to as disaggregated Q10 (DQ10). NDQ10 and DQ10 were found to have a large activation barrier preventing these two conformations from interconverting.[82] Disaggregated peptides, including DQ10, are more resistant to aggregation and fibrillization compared to non-disaggregated peptides.[31,46,57]

In this study we use UVRR spectroscopy, circular dichroism (CD), and metadynamics simulations to investigate the solution-state structures of $D_2Q_{15}K_2$ (Q15) and $D_2Q_{20}K_2$ (Q20) in their disaggregated (D) and non-disaggregated (ND) forms. We compare our results with those previously published for Q10[82] and Q20[193] to investigate the dependence of the solution-state polyQ structures on repeat length. We find that disaggregated Q15 (DQ15) has predominately the same PPII-like structure as previously found for DQ10[82] and disaggregated Q20 (DQ20).[193] Also, non-disaggregated Q15 (NDQ15) has predominately the same collapsed β-strand-like conformation as previously found for NDQ10.[82] Unlike NDQ10 and NDQ15, non-disaggregated

Q20 (NDQ20) is insoluble in water. The structures of NDQ20 were previously described in detail.[193]

Using CD spectroscopy, we show that disaggregated polyQ (DQ) peptides, while predominantly in a PPII-like conformation, have a significant population of collapsed β-strand-like conformation that increases with polyQ length. Our metadynamics simulations show that the relative energies of the PPII-like structures increase and the relative activation energy of the PPII → β-strand conformational transition decreases with increasing polyQ length. From these data we conclude that longer polyQ peptides have an increased preference for the aggregation-prone collapsed β-strand-like conformation compared to the aggregation-resistant PPII-like conformation. This structural preference may in part explain why proteins containing longer polyQ tracts have an increased aggregation rate, which is associated with a decrease in the neurodegenerative disease age of onset.

8.2 Materials and Methods

8.2.1 Materials

The peptides Q15 and Q20 were purchased from Thermo Fisher Scientific at ≥ 95% purity. Trifluoroacetic acid (TFA) was purchased from Thermo Fisher Scientific at ≥ 99.5% purity, and 1,1,1,3,3,3,-hexafluoro-2-propanol (HFIP) was purchased from Acros Organics at ~99% purity.

8.2.2 Sample Preparation

NDQ15 was prepared by dissolving the peptide in nanopure water. Solution-state, non-disaggregated polyQ peptides are sensitive to impurities that can nucleate aggregation. NDQ15 was prepared in sterile centrifuge tubes to remove any nucleating centers.

DQ15 and DQ20 were prepared using the disaggregation protocol developed by Chen et al.[31] Briefly, the peptide was dissolved in 1:1 TFA/HFIP by sonication (~15 min) and incubated for ~2 hours at room temperature. The TFA/HFIP solvent was then evaporated with a stream of dry nitrogen and the peptide was dissolved in water to the desired concentration (0.3 mg/ml for UVRR and CD measurements).

8.2.3 UV Resonance Raman Spectroscopy

The UVRR instrumentation used in this work was previously described by Bykov et al.[103] An Infinity Nd:YAG laser (Coherent, Inc.) was used to generate ~204 nm light by Raman-shifting the third harmonic with H_2 gas (30 psi) and using the fifth anti-Stokes line. We generated 197 nm laser excitation using the 4^{th} harmonic of a tunable Ti:sapphire laser (Positive Light). Samples were contained in a spinning Suprasil quartz NMR tube during spectral collection, and a ~165° backscattering angle was used. The scattered light was dispersed in a subtractive double monochromator. A back-thinned, liquid nitrogen-cooled CCD camera (Spec10:400B, Princeton Instruments) with a Lumagen-E coating was used to detect the Raman scattering.

8.2.4 CD Spectroscopy

CD spectra were measured using a Jasco J-710 spectropolarimeter. Samples were placed in a cylindrical fused silica quartz cuvette with a pathlength of 0.02 cm. Spectra were collected with 0.2 nm data intervals and averaged over ~10 scans.

8.3 Computational Methods

8.3.1 General Simulation Details

Metadynamics simulations were used to investigate the conformational energy landscape of Q15 and Q20. Subsequently, molecular dynamics (MD) simulations were performed on low-energy conformations of the Q15 and Q20 landscapes to obtain structural information on the equilibrated structures. Monomeric Q15 and Q20 peptides were placed in periodic boxes of 11665 and 18658 TIP3P[165] water molecules with dimensions of 72×72×72 Å and 84×84×84 Å, respectively. All simulations were performed with the NAMD[107] MD engine, and simulation trajectory data were analyzed using Visual Molecular Dynamics (VMD) 1.9.2[108] with a Tcl scripting interface.

Potential energies were calculated using the CHARMM36 force field.[109] The particle mesh Ewald algorithm[111] (1.0Å grid spacing) was used to calculate system electrostatics. An integration time step of 2 fs was used. The NPT ensemble (constant atom number, pressure, temperature) was used for all simulations with a Langevin thermostat and piston to maintain

temperature and pressure at 300 K and 1.01225 bar respectively.[112] The pair list distance, interaction cutoff, and switch distance were 14.0, 12.0, and 10.0 Å respectively.

8.3.2 Metadynamics Simulations

Well-tempered metadynamics simulations were used to characterize the conformational free energy landscape of the Q15 and Q20 peptides. Metadynamics is a method that enhances the sampling of a standard MD simulation by adding artificial potentials, in the form of Gaussian functions, to conformations that the system previously sampled. This allows the system to escape deep energy wells and to sample a large region of its conformational energy landscape.[116,117]

The initial Q15 and Q20 peptide monomers used in the metadynamics simulations were fully extended peptides with Ramachandran Φ and Ψ angles of 180°. The peptide systems were energy-minimized for 10 000 steps using the conjugate gradient method and then equilibrated for 500 ps. Metadynamics was then performed on the equilibrated system for 400 ns of simulation time.

To visualize the Q15 and Q20 conformational landscapes, we plotted the metadynamic conformations using the root-mean-square deviation (RMSD) for the α-carbon positions of idealized α-helix, β-hairpin, and PPII conformations of Q15 or Q20. These collective variables are similar to that previously used by Punihaole et al.[82] Ideal α-helix, β-hairpin, and PPII collective variable reference structures were created with Molecular Operating Environment[106] software (MOE 2013.10). The maximum RMSDs for metadynamics structures was set to 15 Å and 20 Å for Q15 and Q20 respectively, and this limit was maintained by a 1.0 kcal mol^{-1} $Å^{-1}$ half-harmonic potential constraint. Artificial potentials, in the form of Gaussian functions with a

height of 1.0 kcal mol^{-1} and a width of 0.3 Å for Q15 and 0.4 Å for Q20, were added to the metadynamics simulations every 100 steps. The metadynamics free energy data and collective variable history were processed using Python scripts to identify the landscape positions and structures for conformations found in local energy wells. The scripts and coordinate files used in the metadynamics simulations are available for download.

Metadynamics simulation parameters were chosen to minimize error and ensure adequate sampling of the energy landscape following the work of Laio et al.[195] Also, Laio et al.[195] empirically derived an equation to estimate the error of metadynamics simulations for systems with energy landscapes containing only a few well-defined energy wells. However, our system contains complex energy landscapes with many local energy minima. Thus, the methods developed by Laio et al.[195] are unable to estimate the error in our simulations.

MD simulations (10 ns) were used to obtain equilibrium structural information on the low-energy conformations found in the metadynamics energy landscapes. Ψ and Φ dihedral angles were extracted from simulation trajectories using a Tcl script in VMD.

8.4 Results and Discussion

8.4.1 UVRR Spectroscopy

UVRR spectroscopy, with a ~204 nm excitation wavelength, is in resonance with the $\pi \rightarrow \pi^*$ transitions of amide groups, including the secondary amides of the peptide backbone[194] and the primary amides of the Gln side chains.[57,71,72] Therefore, the UVRR spectra of our polyQ

peptides are dominated by the resonance-enhanced vibrations located on these amide chromophores. This greatly simplifies the vibrational spectrum of these peptides.

As discussed above, Punihaole et al. previously showed that NDQ10 exists in a β-strand-like conformation while DQ10 is in a PPII-like conformation.[82] Also, Jakubek et al. previously showed that DQ20 is also in a PPII-like conformation.[193] Here, we used UVRR to investigate the structures of solution-state NDQ15 and DQ15.

We previously discussed detailed UVRR band assignments for NDQ10[82], DQ10[82], NDQ20,[193] and DQ20.[193] Here, we find that the UVRR spectra and band assignments of NDQ15 are essentially identical to those of NDQ10.[82] Also, we find that the UVRR spectra and band assignments of DQ15 are essentially identical to those of DQ10[82] and DQ20[193] (Figure 8.1Figure 7.1). As such, we will only discuss the assignment of the conformationally sensitive amide III_3 band of the peptide backbone ($AmIII_3^S$). A complete discussion of assignments can be found in previous work by Punihaole et al.[82]

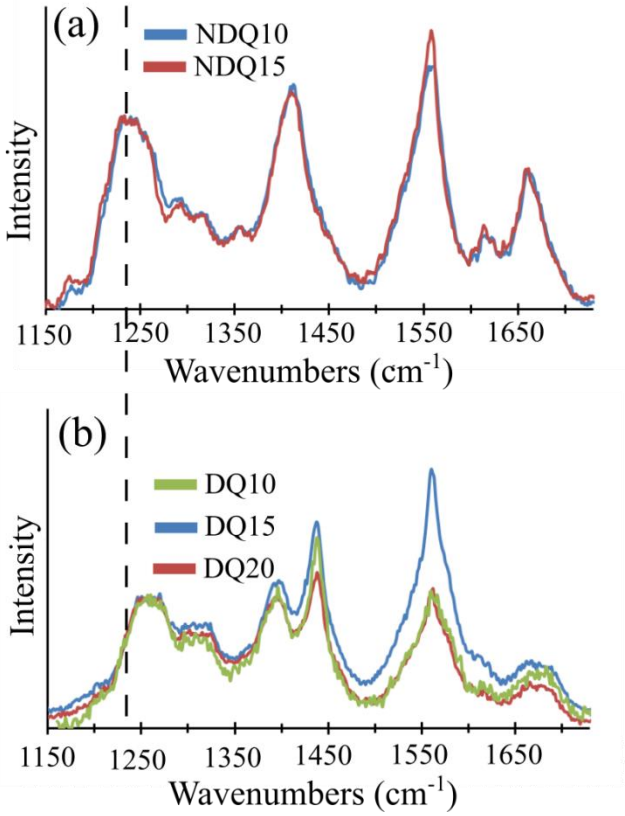

Figure 8.1: UVRR spectra (204 nm) of solution state polyQ peptides: (a) NDQ10 and NDQ15; (b) DQ10, DQ15, and DQ20. All spectra were normalized to the $AmIII_3^S$ band. The dashed line is located at ~1240 cm^{-1}. The UVRR spectra of NDQ10 and DQ10 were previously reported by Punihaole et al.,[82] and the UVRR spectrum of DQ20 was previously reported by Jakubek et al.[193] Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

8.4.2 Ramachandran Ψ Angle Distributions of Q15 and Q20

Asher et al. previously showed that the frequency of the $AmIII_3^S$ band sinusoidally depends on the Ramachandran Ψ angle of the peptide backbone.[65] This correlation results from coupling of the backbone N-H bending motion of the $AmIII_3^S$ band with the $C_\alpha H$ bending vibration at ~1390 cm^{-1}.[65] Mikhonin et al. later developed a method to quantitatively calculate the Ramachandran Ψ angle of a peptide backbone in a variety of different solvation states and sample temperatures.[64]

Asher et al. showed that the Ψ angle distribution of a peptide can be estimated from the inhomogeneous line width of the $AmIII_3^S$ band.[74] Assuming that the inhomogeneous broadening is solely due to the Ψ angle distribution of the peptide, one can model the $AmIII_3^S$ band as a sum of Lorentzian bands with a width equal to the homogenous linewidth of the $AmIII_3^S$ band. Using the equations derived by Mikhonin et al.[64], the corresponding Ψ angle can be calculated for each Lorentzian band producing a Ψ angle distribution. These methodologies have previously been used to estimate the Ψ angle distribution in a variety of peptides[62,76,78] including those of polyQ.[46,57,82,193]

Here, we use these methods to calculate the Ψ angle distributions of NDQ15 and DQ15 (Figure 8.2a-c) from the UVRR spectral fits (Figure 8.3). The UVRR spectrum of NDQ15 contains an $AmIII_3^S$ band centered at ~1240 cm^{-1}. This corresponds to a Ψ angle distribution that peaks at Ψ~140°, which is consistent with β-strand conformations.[196] The $AmIII_3^S$ band and Ψ angle distribution of NDQ15 are essentially identical to that of NDQ10.[82] Thus, we conclude that the secondary structure of NDQ15 is the same as that of the NDQ10 collapsed β-strand-like conformation.[82]

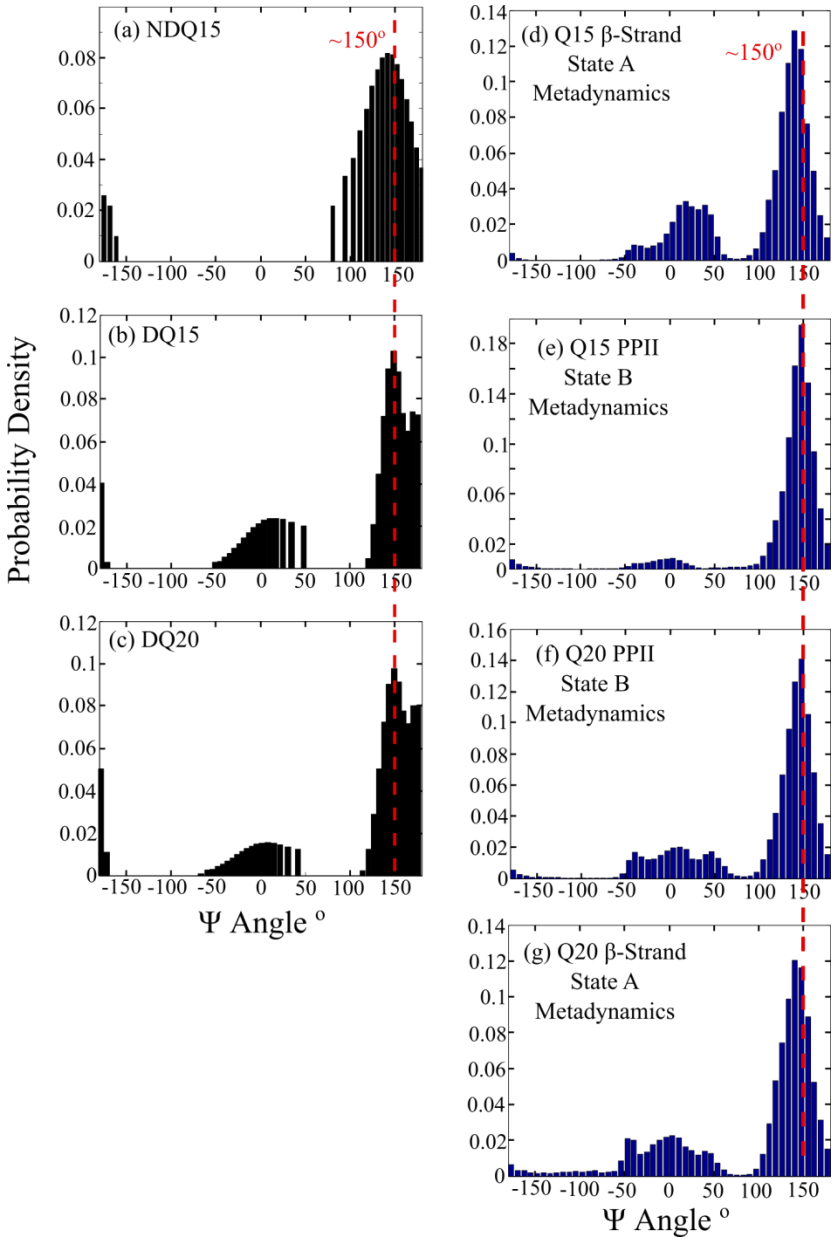

Figure 8.2: Experimental and computational Ψ angle distributions. Experimental Ψ angle distributions of solution-state (a) NDQ15, (b) DQ15, and (c) DQ20. Computational Ψ angle distributions from metadynamics energy wells for (d) Q15 state A, (e) Q15 state B, (f) Q20 state B, and (g) Q20 state A. The experimental and computational Ψ angle distributions of NDQ10 and DQ10 can be found in reference [82]. The experimental Ψ angle distribution for DQ20 was previously reported by Jakubek et al.[193] Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

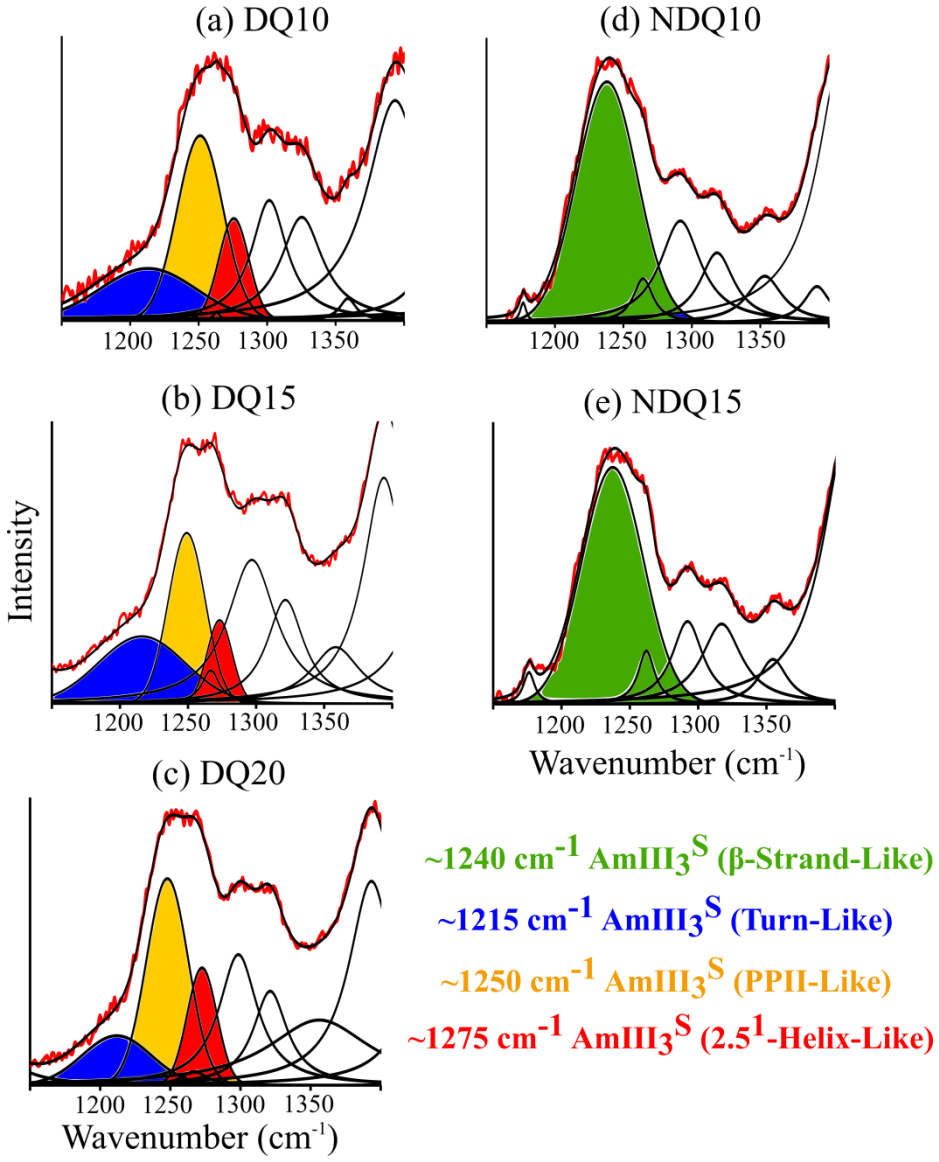

Figure 8.3: Spectral fits for the $AmIII_3^S$ UVRR spectral regions of (a) DQ10, (b) DQ15, (c) DQ20, (d) NDQ10, and (e) NDQ15. The $AmIII_3^S$ bands are color-coded as described in the figure legend. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

DQ15 contains $AmIII_3^S$ bands at ~1275 cm^{-1}, ~1250 cm^{-1}, and ~1215 cm^{-1}, which correspond to Ψ angle distributions peaked at ~175°, ~150°, and ~10°, respectively. Ramachandran Ψ angles of ~175°, ~150°, and ~10° are characteristic of 2.5_1-helix, PPII-helix, and turn-like structures, respectively.[64] The $AmIII_3^S$ bands and Ψ angle distribution of DQ15 are the same as those previously observed for DQ10[82] and DQ20.[193] Thus, we conclude that the secondary structure of DQ15 is the same as that of DQ10/DQ20.[82] From UVRR and metadynamics simulations, DQ10 was previously found to be in a PPII-helix-like conformation interspersed with turn-like structures and with a subpopulation of 2.5_1-helix conformation localized on the charged terminal residues.[82] Additional information regarding our Ψ angle distribution calculations and spectral fitting can be found in Appendix D.

Assuming equal Raman cross sections, we roughly estimate the relative fraction of each secondary structure found in DQ10, DQ15, and DQ20 from the $AmIII_3^S$ band areas. We find that DQ10, DQ15, and DQ20 contain roughly ~55% PPII-like, ~30% turn-like, and ~15% 2.5_1-helix-like conformations. 2.5_1-helix-like conformation (~15%) corresponds to ~2-3 peptide bonds per peptide.

Punihaole et al.[82] previously showed that the 2.5_1-helix structure of DQ10 is localized on the charged terminal residues. This is because the 2.5_1-helix conformation derives primarily from the electrostatic repulsions of adjacent charged amino acids.[119,133] Our observation that DQ10, DQ15, and DQ20 contain ~2-3 peptide bonds in the 2.5_1-helix conformation is consistent with the localization of the 2.5_1-helix on the charged terminal residues.

8.4.3 Metadynamics Structures of Q15 and Q20

To obtain further insights into the solution-state structures of Q15 and Q20, we performed metadynamics simulations to investigate their conformational energy landscapes. The use of metadynamics to examine the polyQ energy landscape was previously demonstrated by Punihaole et al.[82] The metadynamics conformational landscapes of Q15 and Q20 are found in Figure 8.4. As previously observed for Q10, the energy landscapes for Q15 and Q20 are frustrated with many shallow local minima. For both Q15 and Q20 we find two deep, local energy wells on the energy landscape that correspond to different low-energy structures. We will designate these conformations as state A and state B. Representative structures from each energy well are shown in Figure 8.4. We examine the Φ and Ψ angle distributions of conformations found in state A and state B for both Q15 and Q20 (Figure 8.5). The state A and state B conformations are similar for both Q15 and Q20.

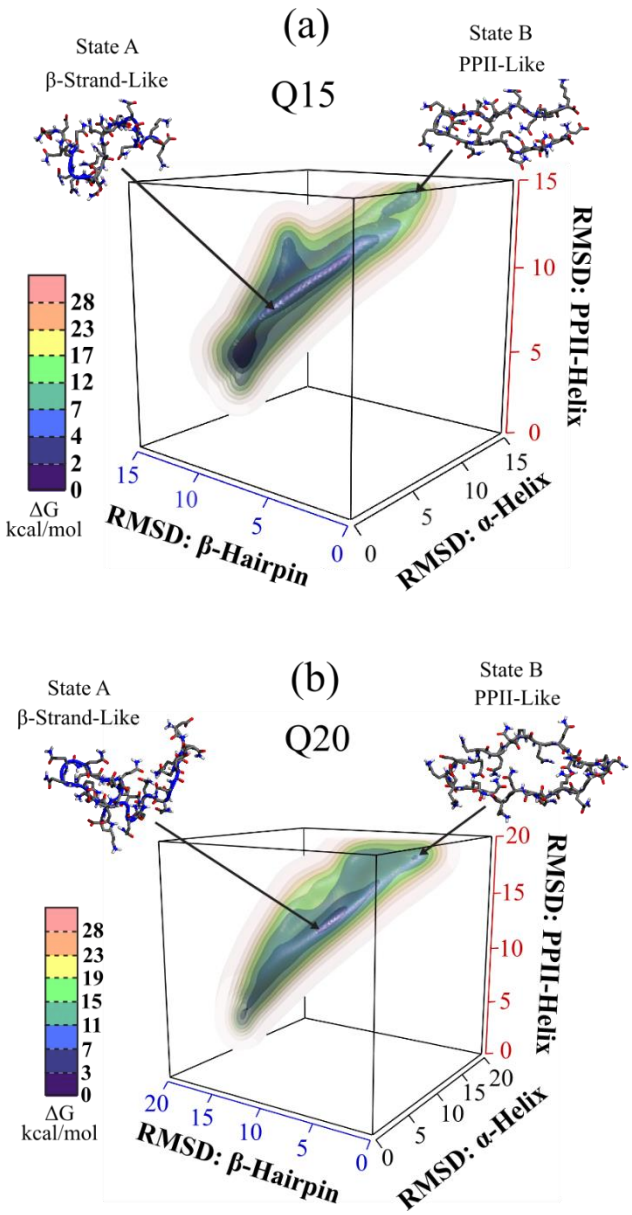

Figure 8.4: Conformational energy landscapes of (a) Q15 and (b) Q20 obtained from metadynamics simulations. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203.

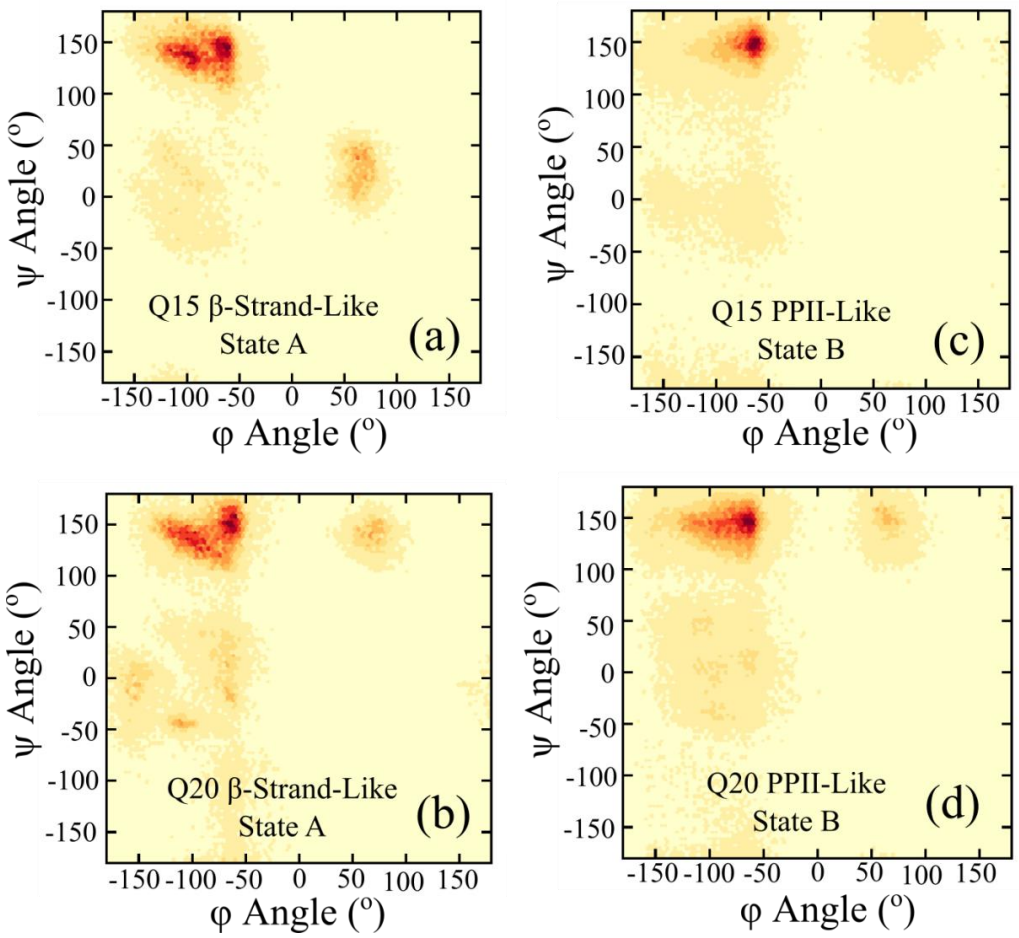

Figure 8.5: Ramachandran plots for (a) Q15 state A, (b) Q20 state A, (c) Q15 state B, and (d) Q20 state B metadynamics structures. Darker red colors indicate an increased number of peptide bonds in a given region of the Ramachandran plot. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society..

State A of Q15 and Q20 both contain peptide bonds with (Φ, Ψ) angle populations of (-95°,140°) and (-65°,140°), which are found in the β-sheet region of the Ramachandran plot.[196] Also, the structures of state A contain a broad Ψ angle distribution centered at ~0°, which is characteristic of turn-like structures.[196] The Ψ angle distributions of the state A structures for Q15 and Q20 are centered at ~140°. This Ψ angle distribution is in quantitative agreement with that observed for NDQ15. Thus we conclude that state A of the metadynamics landscape corresponds to the structure of the non-disaggregated peptide. The structure and (Φ, Ψ) distributions of state A for Q15 and Q20 are similar to the state A metadynamics structure of Q10 previously reported.[82] From our metadynamics simulations we conclude that state A consists of a predominately β-strand-like secondary structure with turn-like structural regions that allow the structure to be collapsed.

For state B structures of Q15 and Q20, we find that the (Φ, Ψ) angle distribution peaks at (-65, 150°). These angles are found in the PPII region of the Ramachandran plot.[197] We also find a significant population of Ψ angles centered at ~0° (see Figure 8.2), which is characteristic of turn-like conformations.[196] The Ψ angle distribution of state B is centered at ~140° (see Figure 8.2). This Ψ angle distribution is in quantitative agreement with those observed for DQ15 and DQ20 (Figure 8.2). Thus we conclude that state B of the metadynamics landscape corresponds to the structures of DQ15 and DQ20. The structure and (Φ, Ψ) distributions of state B for Q15 and Q20 are similar to the state B metadynamics structure of Q10 previously reported.[82] From our metadynamics simulations we conclude that state B consists of predominately PPII-like secondary structures with turn-like structures enabling the structure to be collapsed.

In the experimental Ψ angle distributions calculated from UVRR spectra of DQ10-20 we find a peak at ~175° which we assign to 2.5_1-helix conformations. However, we do not observe

this peak in the state B metadynamics structure. As discussed previously by Punihaole et al.[82] and Feng[135], this discrepancy likely results from an inadequacy in the CHARMM36 force field. The CHARMM36 force field was optimized using X-ray crystal structures of proteins.[134] The 2.5_1-helix is an uncommon structure that is found only in peptide sequences with adjacent charged amino acids.[133] Thus, it is unlikely that this structure is well parameterized in the CHARMM36 force field.

For Q15 and Q20, the β-strand-like energy well (state A) in the metadynamics simulations is conformationally broad. 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 with significant conformational variability. In contrast, the PPII-like energy well (state B) of Q15 and Q20 is narrow, indicating a more well-defined structure with less structural variability.

In contrast, metadynamics simulations of Q10 have revealed narrow energy distributions for both the PPII-like and β-strand-like structures.[82] Our observation that the Q10 peptide contains less conformational variability compared to Q15 and Q20 is in agreement with the computational work of Wang et al.[37] They used molecular dynamics simulations to investigate the structures of Q5 and Q15 peptides. In contrast to this study, their results indicate that both Q5 and Q15 are disordered in solution. However, they find that Q15 has increased conformational variability compared to Q5. This is thought to result from an increase in the number of possible peptide-peptide hydrogen bonding contacts with longer polyQ tracts. They conclude that longer polyQ peptides have increased aggregation kinetics because their conformational variability allows the peptide to sample β-sheet structures that nucleate fibrillization.[37]

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.[82]

8.4.4 Circular Dichroism Spectroscopy of Q15 and Q20

8.4.4.1 CD Spectra of DQ15, DQ20, and NDQ15

To further investigate the structures of Q15 and Q20, we collected CD spectra of NDQ15, DQ15, and DQ20 (Figure 8.6). The CD spectrum of NDQ15 consists of a negative peak at ~218 nm and a strong positive peak at ~196 nm that are characteristic of β-sheet conformations.[198] In contrast, the spectra of both DQ15 and DQ20 have a strong negative peaks at ~200-205 nm, which are characteristic of PPII conformations.[198] Our CD spectra are consistent with our UVRR measurements and metadynamics simulations, concluding that our DQ peptides are in a PPII-like conformation and our NDQ peptides are in a β-strand-like conformation.

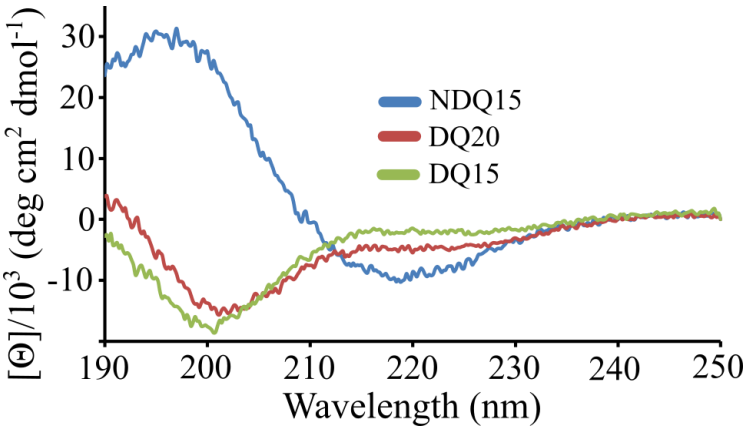

Figure 8.6: CD spectra of NDQ15 (blue), DQ15 (green), and DQ20 (red). Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

8.4.4.2 DQ20 and DQ15 Have a Population of β-Strand Structures

We compared the CD spectra of DQ15, DQ20, and DQ10 (Figure 8.7a). The CD spectrum of DQ10 was previously measured by Punihaole et al.[82] Both the DQ20 and DQ15 spectra show a less negative and slightly red-shifted peak at ~200-205 nm and a more positive peak at ~190 nm compared to DQ10. Also, the trough at ~220 nm is more negative for DQ20 compared to that for DQ10. To highlight the spectral differences, we subtracted the DQ10 spectrum from that of DQ20 and DQ15 (Figure 8.7b). The difference spectra contain a negative weak peak at ~220 nm and a stronger positive peak at ~195 nm. This is characteristic of β-sheet conformations[198] and is similar to the CD spectra of NDQ10 and NDQ15.

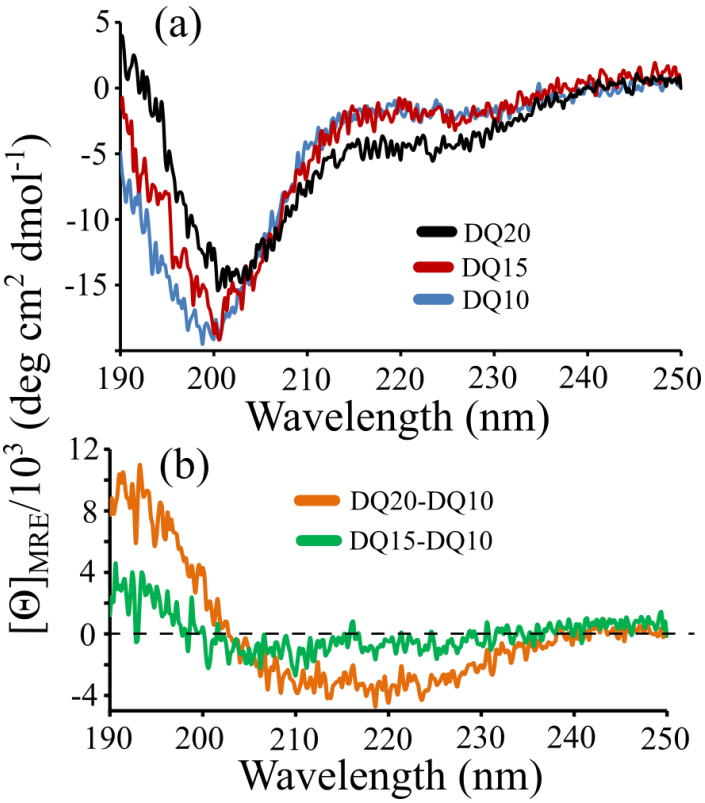

Figure 8.7: Comparison of DQ10-20 CD spectra. (a) CD spectra of (blue) DQ10, (red) DQ15, and (black) DQ20. (b) CD difference spectra of (orange) DQ20-DQ10 and (green) DQ15-DQ10. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

We modeled the spectra of DQ15 and DQ20 as a linear combination of DQ10 and NDQ15 basis spectra. We chose these basis spectra because these peptides best represent the pure β-strand-like and PPII-like conformations of polyQ, respectively (see Appendix D for details). The spectra of DQ15 and DQ20 are well modeled using NDQ15 and DQ10 CD spectra as basis spectra (Figure D.1 and Figure D.2). This suggests that the CD spectral differences between DQ10, DQ15, and DQ20 result from differing amounts of a minority β-strand-like population in the DQ peptides. We used our model to estimate the percentage of collapsed β-strand-like structures in DQ15 and DQ20 (see Appendix D). We find that DQ15 and DQ20 contain ~5% and ~15% β-strand conformation respectively. These data show that, at least for Q10-Q20, longer DQ peptides contain an increasing population of the β-strand-like structure.

Our result is in great agreement with that reported by Chellgren et al. who examined the CD spectra of disaggregated polyQ peptides containing 1-15 Gln residues.[98] They find that the negative peak at ~205 nm decreases in magnitude and the positive peak at ~225 nm increases in magnitude with increasing polyQ tract length, which is similar to what we observe for DQ10-20. From their CD data, Chellgren et al. conclude that longer DQ peptides have decreased PPII content and increased β-strand/sheet content.[98] They report that their DQ15 peptide has ~7% β-strand population, which is similar to what we observe for DQ15.

We also compared the CD spectra of NDQ15 and NDQ10 (Figure 8.8). The CD spectrum of NDQ10 was previously measured by Punihaole et al.[82] We find that the CD spectra of NDQ10 and NDQ15 are essentially identical. This indicates that the NDQ10 and NDQ15 peptides do not have a structural dependence on polyQ length.

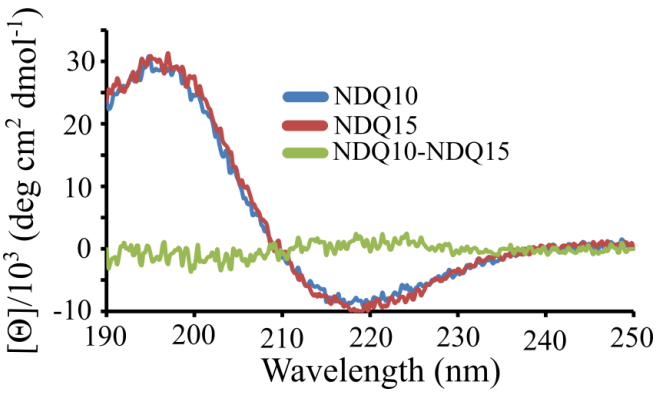

Figure 8.8: CD spectra of (blue) NDQ10, (red) NDQ15, and (green) NDQ10-NDQ15 difference spectrum. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

8.4.4.3 Detectivity of UVRR Spectroscopy to Minority PPII-Like and β-strand-like Populations

Our CD measurements show that DQ15 and DQ20 contain minority populations of β-strand-like structures. However, these subpopulations are not clearly observed in our UVRR spectra. One possible origin for this discrepancy is that the UVRR spectra may not be adequately sensitive to the small subpopulation of β-strand-like secondary structures in the presence of a dominantly PPII-like conformation.

We examined the sensitivity of 204 nm excited UVRR spectroscopy to detect β-strand-like subpopulations in DQ peptides by artificially adding a β-strand-like $AmIII_3^S$ peak (1240 cm^{-1}, 40 cm^{-1} FWHH) to the spectral fit of DQ20. The area of the 1250 cm^{-1} PPII-like $AmIII_3^S$ band was reduced to conserve the total $AmIII_3^S$ band area. This spectrally models an increase in the β-

strand-like population in the presence of a corresponding decrease in the PPII-like conformation, assuming equal Raman cross sections.

We qualitatively compared the UVRR spectrum of DQ20 to spectral fits that model 10%, 20%, 30%, 50%, and 80% β-strand-like minority populations. We find that the spectral fit does not significantly differ from that of the DQ20 spectrum until the β-strand subpopulation exceeds ~20%. We conclude that the UVRR spectra have low sensitivity to β-strand-like subpopulations because the 1240 cm^{-1} β-strand-like $AmIII_3^S$ band is broad (40 cm^{-1} FWHH) and overlaps more intense bands in the $AmIII_3^S$ spectral region. Our result suggests that the UVRR limit of detection for β-strand subpopulations is ~≥20%.

8.4.5 Gibbs Free Energies and Activation Energies for PolyQ Metadynamics Structures

The Gibbs free energies and activation barriers for the metadynamics polyQ structures are summarized in Figure 8.9.

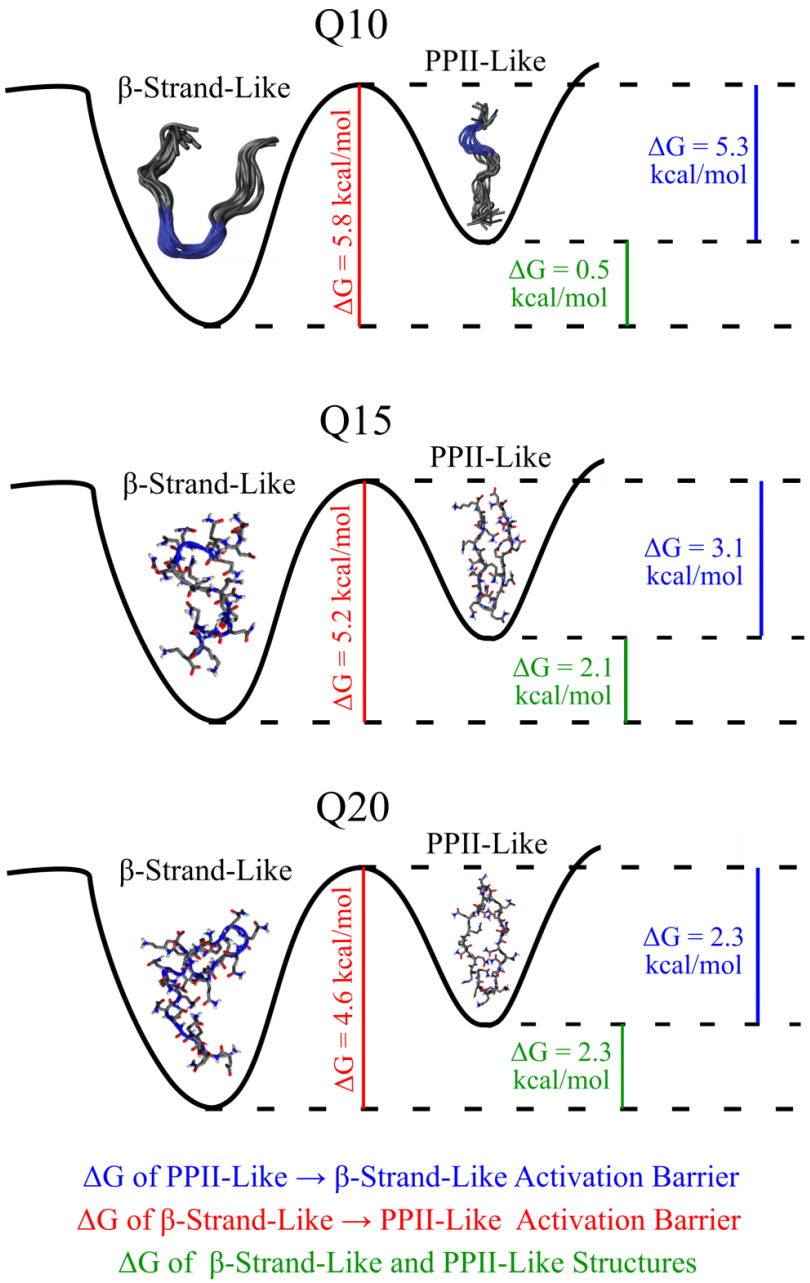

Figure 8.9: Depiction of relative energies and activation barriers for the Q10-20 PPII-like and β-strand-like conformations. The activation barrier ΔG for the β-strand→PPII transitions are shown in red, the activation barrier ΔG for the PPII→β-strand transitions are shown in blue, and the relative ΔG of the β-strand-like and PPII-like minima are shown in green. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

8.4.5.1 Gibbs Free Energies of the β-Strand-Like and PPII-Like Metadynamics Structures

Table 8.1 shows the difference in the Gibbs free energy of the β-strand-like and the PPII-like conformations calculated for Q10, Q15, and Q20 using metadynamics. The β-strand-like conformation is found at the global energy minimum of the landscape for Q10, Q15, and Q20. We find that the difference in the Gibbs free energy between the PPII-like and β-strand-like conformations increases as the polyQ tract length increases. From this result, we conclude that the PPII-like conformation is less energetically favorable, compared to the β-strand-like conformation, for longer polyQ peptides.

Table 8.1: Difference in the Gibbs Free Energies of the β-Strand-Like and PPII-Like Conformations of Q10, Q15, and Q20. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

	(ΔG PPII)-(ΔG β-strand) (kcal/mol)
Q10	0.5
Q15	2.1
Q20	2.3

8.4.5.2 Activation Barriers between PPII and β-Strand Conformations

Using metadynamics, we calculated the activation barriers for the PPII-like→β-strand-like and β-strand-like→PPII-like conformational transitions of Q10-20. As stated above, the energy landscapes of Q10-20 are rough with many local minima. Thus, the activation barrier for

interconversion between structures heavily depends on the pathway the peptide takes through its conformational landscape.

We calculated the lowest energy activation barriers for the PPII-like→β-strand-like and β-strand-like→PPII-like structural conversions from the metadynamics free energy landscapes. The lowest energy pathway between the PPII-like and β-strand-like energy wells was identified using a Monte Carlo method, where random pathways were sampled and their activation barriers minimized. The lowest activation barrier between the PPII-like and β-strand-like states was identified, as the Monte Carlo simulations converged upon a pathway with minimum activation energy. The values reported in Table 8.2 are those of the lowest energy pathways for the respective structural conversion.

Table 8.2: Activation Barrier Energies for the β-strand-like → PPII-like and PPII-like → β-strand-like Conformational Transitions in Q10, Q15, and Q20. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

	β-Strand → PPII (kcal/mol)	PPII → β-Strand (kcal/mol)	(β-Strand→PPII) – (PPII→β-strand) (kcal/mol)
Q10	5.8	5.3	0.5
Q15	5.2	3.1	2.1
Q20	4.6	2.3	2.3

From the metadynamics simulations, we find that the PPII→β-strand activation barriers are lower compared to the β-strand→PPII activation barriers for Q10, Q15, and Q20. We observe a decrease in both the β-strand→PPII and PPII→β-strand energy barriers with increasing peptide length. Also, we examined the energy difference between the β-strand→PPII and PPII→β-strand activation barriers. We find that the PPII→β-strand activation barrier decreases relative to the β-strand→PPII activation energy with increasing polyQ length.

Our metadynamics results for the relative energies and activation energies of Q15 and Q20 indicate that longer polyQ peptides increasingly prefer the β-strand-like conformation. This is in agreement with our CD data showing that, for DQ10-20, longer polyQ peptides have an increased population of β-strand-like conformation. Overall, these data show that the aggregation-prone β-strand-like conformation is increasingly preferred for longer polyQ peptides.

It is possible that increased hydrogen bonding contacts for larger polyQ peptides stabilizes the β-strand-like structure over the PPII-like structure. Walters et al.[44] used fluorescence resonance energy transfer (FRET) to show that longer polyQ peptides are more collapsed. They hypothesized that longer polyQ peptides have increased possible peptide-peptide hydrogen bonding contacts resulting in an unordered collapsed structure.[44] PPII structures are stabilized by water-peptide interactions[182–186] while β-strand structures are stabilized by peptide-peptide hydrogen bonding.[21,193] Therefore, we hypothesize that longer polyQ peptides may have increased hydrogen bonding contacts that stabilize the β-strand-like conformation over the PPII-like conformation.

In addition, Walters et al.[44] hypothesized that the increased number of possible peptide-peptide hydrogen bonding contacts for longer polyQ peptides increases the probability of their formation. Jakubek et al.[193] previously proposed that the PPII→β-strand activation energy is associated with the disruption of peptide-water interactions while forming intrapeptide hydrogen bonds. Thus, it is possible that the increased probability of longer polyQ tracts forming intrapeptide hydrogen bonds results in a decreased PPII→β-strand activation energy.

8.4.5.3 Q15-20 Conformation and Aggregation Kinetics

Our CD results show that, for DQ10-20 peptides, longer polyQ peptides have an increased population of collapsed β-strand structure. Also, our metadynamics simulations show that the β-strand conformation is increasingly thermodynamically favored for longer polyQ peptides. These results are important because the disaggregated PPII-like conformation resists aggregation, while the non-disaggregated collapsed β-strand-like conformation readily

aggregates.[46,57] Thus, the increase of collapsed β-strand population of longer DQ peptides may contribute to their increased aggregation kinetics.[192]

Our result is similar to that of Darnell et al. who used CD spectroscopy to examine the structure of polyQ peptides of different lengths.[199] They found that short polyQ peptides are found in a PPII-like conformation and that increasing the length of the polyQ tract decreases the PPII-like content and increases the β-sheet content of polyQ. From this, they conclude that polyQ peptides are in a "tug-of-war" between PPII and β-sheet structures where longer polyQ peptides prefer β-sheet structures, including fibrils. Our results provide further evidence that longer DQ peptides prefer β-strand structures.

8.4.6 Temperature Dependence of the NDQ15 and DQ20 CD Spectra

We also investigated the temperature dependence of the NDQ15 and DQ20 CD spectra (Figure 8.10). For NDQ15 we find that both the positive peak at ~195 nm and the negative peak at ~220 nm decrease in magnitude with increasing temperature. This spectral change is indicative of a decrease in the β-strand-like population of NDQ15 with increasing temperature. In agreement, the 20-50°C difference spectrum shows a positive peak at ~195 nm and negative peak at ~220 nm which is characteristic of β-sheet conformations. Also, this decrease in β-strand-like conformation is reversible upon cooling the sample from 50°C back to 20°C (Figure 8.10a).

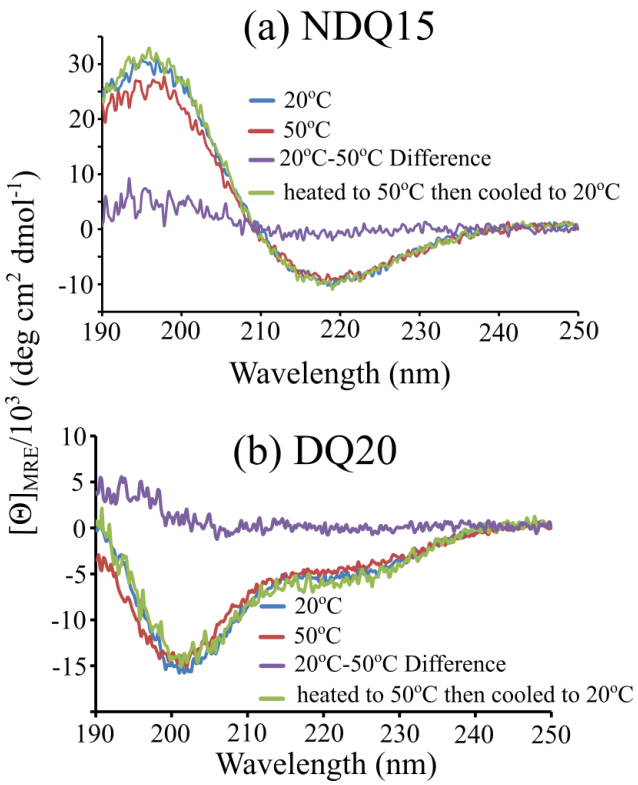

Figure 8.10: Temperature dependence of the (a) NDQ15 CD spectra and (b) DQ20 CD spectra. The blue spectra were collected at 20°C, the red spectra were collected at 50°C, and the purple spectra are the 20-50°C difference spectra. The green spectra were collected after the sample was heated to 50°C and cooled back down to 20°C. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

We find that the CD spectrum of DQ20 is also sensitive to temperature. Upon increasing the temperature of DQ20 from 20 to 50°C we observe a decrease in intensity at ~195 nm and a decrease in the trough at ~220 nm. These spectral changes are consistent with a decrease in the β-strand-like population of DQ20 with increasing temperature. The DQ20 20-50°C difference spectrum shows a peak at ~195 nm which is consistent with our conclusion that the lower-temperature DQ20 sample contains a larger population of β-strand-like structure (Figure 8.10b).

Our result is similar to that of Bhattacharyya et al. who reports the 5-35°C difference spectrum of disaggregated $K_2Q_{40}K_2$ (Q40).[40] They find that the difference spectrum contains a moderate negative peak at ~220 nm, a very weak negative peak at ~205 nm, and a strong positive peak at ~195 nm. This is similar to the CD spectrum we observe for the β-strand-like NDQ10 and NDQ15 peptides. However, Bhattacharyya et al. conclude that this difference spectrum is characteristic of α-helices and that DQ40 forms α-helix conformations at low temperatures.[40]

Bhattacharyya et al. may conflate the β-strand signature of the 5-35°C difference spectrum with that of an α-helix, which has a similar CD spectrum.[40] This may be in part due to the presence of a very weak negative peak at ~205 nm in their 5-35°C difference spectra. The α-helix spectrum consists of a strong positive peak at ~190 nm, a moderate negative peak at ~208 nm, and a weak negative peak at ~222 nm.[198] Thus, the presence of a peak at ~205 nm may suggest α-helix conformation. However, this negative peak could also arise from the incomplete subtraction of the PPII structure, which has a strong negative peak at ~205 nm, in the 5°C-35°C difference spectrum.

8.4.7 Non-disaggregated PolyQ Conformational Flexibility May Contribute to Fast Aggregation Kinetics

As discussed above, we find that the β-strand-like energy wells in the metadynamics simulations are broad compared to the PPII-like energy wells for Q15 and Q20. Thus, the β-strand-like conformation can more easily sample different structures compared to the PPII-like conformation. This agrees with the temperature dependence of the NDQ15 and DQ20 CD spectra that, in both cases, show a decrease in β-strand population with increasing temperature.

In contrast, the metadynamics PPII-like energy wells are narrow, indicating that the PPII-like structure cannot easily sample different conformations. This is in agreement with the temperature dependence of the DQ20 CD spectrum that indicates no change in the PPII-like structure with increasing temperature. Also, the CD spectrum of DQ15 shows less temperature sensitivity compared to that of DQ20 (Figure D.4). This is a result of DQ15 containing a smaller subpopulation of the temperature sensitive β-strand-like conformation and a larger, temperature-insensitive PPII-like population.

Previously, Vitalis et al. and Wang et al. proposed that the conformational flexibility of polyQ peptides promotes their aggregation.[37,38] The flexibility of polyQ is thought to promote aggregation by increasing intermolecular interactions and/or increasing the probability of sampling β-sheet conformations that are prone to fibrillization. Both our metadynamics simulations and CD spectra show evidence that the β-strand-like conformation of NDQ peptides has more conformational freedom than the PPII-like structure of DQ peptides. In addition, it is well known that NDQ peptides have much faster aggregation kinetics compared to DQ

peptides.[31,46,82] Thus, increased conformational flexibility of the β-strand-like conformation may contribute to the fast aggregation rate of NDQ peptides.

8.5 Conclusions

Here, we use UVRR spectroscopy, CD spectroscopy, and metadynamics simulations to investigate the structures of DQ15, NDQ15, and DQ20. We find that NDQ15 exists in predominately a β-strand-like conformation while DQ15 and DQ20 exist in predominately a PPII-like conformation. These results are similar to those previously reported for DQ10 and NDQ10.[82]

We compare the CD spectra of DQ10, NDQ10, DQ15, NDQ15, and DQ20. We find that DQ peptides contain a minority population of the collapsed β-strand-like conformer. We show that longer DQ peptides have larger populations of the collapsed β-strand structure. Our metadynamics simulations show that the PPII-like conformation becomes relatively less energetically favorable and that the PPII → β-strand conformational transition becomes relatively more energetically favorable with increasing polyQ length. This indicates that longer polyQ peptides have an increased propensity for the aggregation-prone collapsed β-strand-like conformation, compared to the aggregation-resistant PPII-like conformation. Because the β-strand-like conformation is more prone to aggregation, the increased β-strand-like conformational propensity in longer polyQ peptides may significantly increase the aggregation kinetics of longer polyQ tracts.

We also used CD spectroscopy to investigate the effects of temperature on the structure of NDQ15 and DQ20. We found that increasing temperature decreases the population of the β-

strand-like conformations in both NDQ15 and DQ20. However, we observe no changes in the PPII-like content of DQ10-20 peptides with increasing temperature. Metadynamics simulations show that the β-strand-like energy well is broad allowing for structural freedom, while that of the PPII-like structure is narrow suggesting a more well-defined structure with less conformational flexibility. Computational work previously showed that increased conformational flexibility promoted aggregation and fibrillization of polyQ peptides.[37,38] Thus, conformational flexibility may play a role in the fast aggregation kinetics of NDQ peptides in β-strand-like conformations.

Overall, our work provides crucial insights into the connection between polyQ peptide length, structure, and aggregation kinetics. We find that the β-strand-like conformation of polyQ has more conformational freedom compared to the PPII-like conformation, which may contribute to the faster aggregation kinetics of the β-strand-like structure. Our work also shows that longer polyQ peptides increasingly prefer the aggregation-prone β-strand-like conformation over the aggregation-resistant PPII-like conformation. This structural preference likely plays a role in increasing the aggregation kinetics of longer polyQ peptides, which is associated with an earlier age of onset for polyQ neurodegenerative diseases.

8.6 Author Information

Present Addresses: R.J.W.: Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas 77555-0304, United States

8.7 Acknowledgements

Funding for this work was provided by the University of Pittsburgh (R.S.J., S.E.W., S.A.A.), the Defense Threat Reduction Agency HDTRA-09-14-FRCWMD (R.S.J., S.A.A.), and was partially supported by NIH R01 DA027806 (R.J.W.). S.E.W. gratefully acknowledges support through the Molecular Biophysics and Structural Biology NIH Training Grant (T32 GM 088119). The MD simulation computer time was supported by XSEDE MCB060069, and computer equipment was purchased from NSF funds (CHE01126465 and P116Z080180) (R.J.W.).

8.8 Supporting Information

Supporting Information Can be found in Appendix D and includes the following: additional information on UVRR spectral fitting, Ψ angle calculations, modeling of DQ15-20 CD spectra, and UVRR modeling of β-strand sub-populations in DQ peptides. Coordinate files for the α-helix, β-strand, and PPII collective variables as well as python scripts used in the metadynamics simulations are available for download.

9.0 Structure and Hydrogen Bonding Interactions of Non-Fibrillar Polyglutamine Oligomers

Authors: Ryan S. Jakubek, Krishnan Damodaran, Stephen E. White, and Sanford A. Asher.

Author Contributions: R.S.J. collected, analyzed, and interpreted the Raman data with the assistance of S.E.W. K.D. collected and interpreted the DOSY data with assistance by R.S.J. and S.E.W. The manuscript was prepared by R.S.J. and S.A.A. with assistance by K.D. and S.E.W.

Proteins containing expanded polyglutamine (polyQ) tracts aggregate into fibrils. This aggregation is associated with numerous neurodegenerative diseases including Huntington's disease. Longer polyQ tracts have faster aggregation kinetics, which are associated with earlier age-of-onset for disease patients. Thus, there is great interest in understanding the dependence of aggregation rate on the polyQ tract length. Relevant experimental investigations are impeded by the poor water solubility of polyQ peptides with Q≥20. To circumvent this, polyQ peptides are generally subjected to a disaggregation procedure that increases their apparent solubility. Here, we use diffusion-ordered NMR spectroscopy (DOSY) and UV resonance Raman spectroscopy (UVRR) to investigate the concentrations and hydrogen bonding interactions of solution-state, non-fibrillar polyQ oligomers and their dependence on polyQ repeat length. We examine the small polyQ peptides $D_2Q_{10}K_2$ (Q10), $D_2Q_{15}K_2$ (Q15), and $D_2Q_{20}K_2$ (Q20) in their disaggregated and non-disaggregated forms. We find that disaggregated polyQ peptides are monomeric with negligible inter- and intra-peptide hydrogen bonding regardless of the polyQ tract length. In contrast, non-disaggregated polyQ peptides contain both monomeric and oligomeric species. We

find that longer non-disaggregated polyQ peptides have an increasing oligomer population and that the formation of non-disaggregated polyQ peptide oligomers is driven predominantly by side chain-peptide hydrogen bonding interactions. Also, UVRR data shows that the non-disaggregated polyQ secondary structure is not dependent on the fraction of oligomeric species. These results indicate that non-disaggregated polyQ in monomeric and oligomer forms have the same β-strand-like secondary structure.

9.1 Introduction

Proteins containing expanded polyglutamine (polyQ) tracts are prone to aggregate into fibrils. This fibrillization is associated with at least ten neurodegenerative diseases, including Huntington's disease.[1] Longer polyQ tracts show an increased rate of aggregation and fibrillization[35,192] that is associated with an earlier disease age-of-onset.[97] However, the mechanism of toxicity is still debated and evidence supports both aggregate[14] and monomeric[16,166] polyQ structures as the potential toxic agent.[200]

There is much interest in understanding the mechanism by which longer polyQ peptides have faster aggregation kinetics. However, these studies are limited by the poor aqueous solubility of polyQ peptides with ≥~20 Q residues. To circumvent this, Chen et al. developed a "disaggregation" protocol that solubilizes polyQ peptides.[31] In this method, the polyQ peptide is dissolved in a 1:1 mixture of trifluoroacetic acid (TFA) and hexafluoroisopropanol (HFIP). The peptide is incubated for 0.5-4 hours before the solvent is evaporated under a stream of inert gas. The peptide is then dissolved in water to the desired concentration. This method is commonly referred to as "disaggregation" because it is thought to remove trace aggregates that would seed

fibrillization without changing the secondary structure of the peptide monomer.[33] However, the exact mechanism by which disaggregation solubilizes polyQ peptides is poorly understood. The disaggregated polyQ (DQ) peptide is highly soluble in water and has slower aggregation kinetics compared to the non-disaggregated polyQ (NDQ) species.

Because of the limited aqueous solubility of NDQ peptides with ≥~20 Q residues, most investigations into the structure of polyQ peptides examine the DQ species. Thus, there are limited studies on the structure of solution-state NDQ peptides. The structure of solution-state DQ peptides have been extensively studied using circular dichroism (CD)[18,34,35,98,201] and NMR[18,41,98,201] spectroscopies. Most CD studies indicate that DQ peptides are in a random coil conformation. However, the CD spectrum of unordered and polyproline II (PPII) conformations are similar and difficult to distinguish.[198] Most NMR experiments[18,41,189,201] indicate that DQ peptides are in a random coil conformation, however, the NMR technique used only broadly classifies a peptide into α-helix, β-sheet, or random coil conformations where "random coil" refers to any structure other than α-helices and β-sheets.[202,203] In addition, MD simulations of polyQ peptides indicate a random coil conformation.[37–39,99] However, MD simulations of polyQ peptides are hindered by their rough energy landscape that may prevent the polyQ peptide from sampling structures of its global energy minimum.[82]

In contrast, Chellgren et al.[98] used total correlation spectroscopy (TOCSY) and nuclear Overhauser effect spectroscopy (NOESY) NMR experiments on DQ peptides to show that they contain Φ and Ψ Ramachandran angles consistent with a PPII conformation. From their NMR data, Chellgren et al.[98] also interpreted their CD results as indicating PPII conformations.

Recently, UV resonance Raman (UVRR)spectroscopy was used to investigate the structures of $D_2Q_{10}K_2$ (Q10), $D_2Q_{15}K_2$ (Q15), and $D_2Q_{20}K_2$ (Q20) in their DQ and NDQ

forms.[82,193,204] UVRR spectroscopy provides direct insight into the Ψ angle distribution[64,65] and hydrogen bonding interactions[81] in peptides. They found that DQ10, DQ15, and DQ20 all predominantly have the same PPII-like secondary structure while NDQ10, NDQ15, and NDQ20 all have the same β-strand-like structure. In addition, metadyanmics simulations showed that the conformational energy landscapes of Q10, Q15, and Q20 contain low-energy β-strand-like and PPII-like conformations.[82,204] Using diffusion-ordered NMR spectroscopy (DOSY) Punihaole et al. also showed that the Gln residues of both NDQ10 and DQ10 diffuse with rate constants consistent with monomeric peptides.[82] Thus, they concluded that the PPII-like and β-strand-like structures of DQ10 and NDQ10, respectively, are that of monomeric peptide. These results showed that the disaggregation protocol[31] changes the secondary structure of polyQ peptides from a β-strand-like to a PPII-like conformation. Therefore, a majority of solution-state polyglutamine structural and aggregation studies are specific to polyQ peptides that have had their secondary structures converted to a PPII-like conformation using the disaggregation protocol.

Previously, we used UVRR and other techniques to investigate the structural dependence on the polyQ tract length for Q10, Q15, and Q20. Here we use DOSY and UVRR to investigate the nature of non-fibrillar, solution-state oligomers and their dependence on polyQ peptide length. DOSY NMR is used to examine the presence and quantity of solution-state, non-fibrillar oligomers while UVRR data is used to examine their hydrogen bonding interactions. Here we use a combination of newly reported UVRR data and data originally reported by Jakubek et al.[193,204] and Punihaole et al.,[82] to examine polyQ hydrogen bonding dependence on peptide length.

We find that DQ10-20 peptides are monomeric, with negligible inter- and intra-peptide hydrogen bonding regardless of the polyQ tract length. In contrast, longer solution-state NDQ peptides have an increased population of non-fibrillar oligomers. Also, we find that longer solution-state NDQ peptides have increased side chain-peptide hydrogen bonding, but not backbone-peptide hydrogen bonding, that is correlated to the increased oligomer population. This indicates that the polyQ non-fibrillar oligomers are predominantly stabilized by side chain-peptide hydrogen bonding.

Also, as discussed above, NDQ10-20 species all have the same β-strand-like secondary structure.[204] Here we show that longer NDQ peptides have increased oligomer population. These results indicate that polyQ peptides in non-fibrillar oligomer and monomer forms have the same β-strand-like structure.

9.2 Results and Discussion

Punihaole et al.[46,82] showed that both DQ10 (PPII-like) and NDQ10 (β-strand-like) solution-state conformations aggregate to form fibrils. They found that both solution-state NDQ10 and DQ10 are stable at room temperature and low pH (pH=~2-3), showing no signs of aggregation for >1 week after sample preparation. However, when incubated at ~60°C and ~pH 7, NDQ10 will form fibrils in ~1-2 days[57] and DQ10 will form fibrils in ~1 week.[46,57] Similarly, Jakubek et al.[204] found that solution-state NDQ15, DQ15, and DQ20 peptides show no signs of aggregation for >1 week after sample preparation at room temperature and low pH (pH = ~2-3). Here we study the non-fibrillar, solution-state oligomers of NDQ10, DQ10, NDQ15, DQ15, NDQ20, and DQ20 peptides and their peptide-peptide hydrogen bonding interactions.

9.2.1 Diffusion-Ordered NMR Spectroscopy

We used DOSY NMR spectroscopy to measure the translational diffusion coefficients of DQ15, DQ20, and NDQ15. We examined the signal decay of the lysine (Lys) and glutamine (Gln) side chain methylene proton resonances as a function of gradient strength (see Experimental Methods and Supporting Information for details). The Gln residue diffusion coefficients of NDQ10 and DQ10 were previously reported by Punihaole et al.[82] As described by Jakubek et al.[193], NDQ20 does not dissolve in water. Thus, we were unable to collect DOSY spectra of NDQ20. Figure 9.1 shows the signal decays for the ^{1}H resonances of DQ15, DQ20, and NDQ15. The diffusion coefficients are listed in Table 9.1.

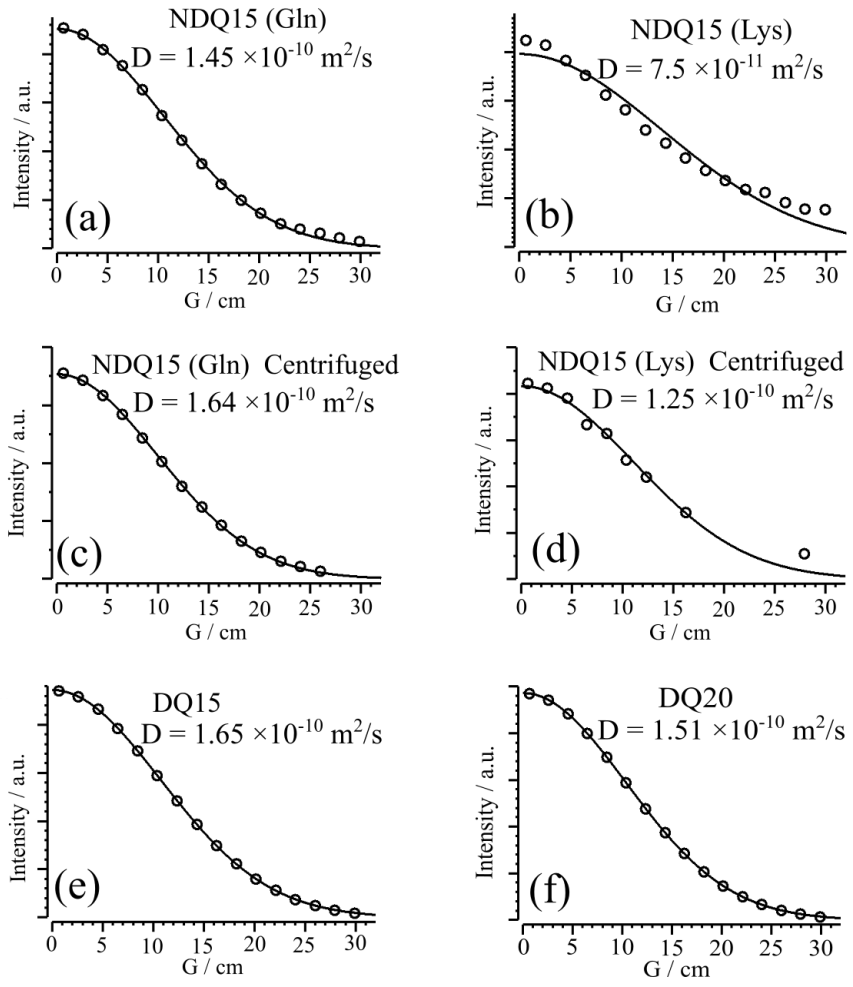

Figure 9.1: DOSY NMR signal decays for (a) NDQ15 Gln residues, (b) NDQ15 Lys residues, (c) NDQ15 Gln residues after centrifugation, (d) NDQ15 Lys residues after centrifugation, (e) DQ15, and (f) DQ20. Data for DQ10 and NDQ10 were previously collected by Punihaole et al. [82] and can be found in Figure E.3.

Table 9.1: Diffusion Coefficients for NDQ and DQ Peptides Measured by DOSY NMR.

	Diffusion Coefficient ($\times 10^{-10}$ m^2/s)	
	Gln ^{1}H Resonance	Lys ^{1}H Resonance
NDQ10	1.94(±0.05) [82]	1.42(±0.05)
NDQ15 Not Centrifuged	1.45(±0.05)	0 .75(±0.05)
NDQ15 Centrifuged	1.64(±0.05)	1.25(±0.05)
DQ10	1.88(±0.05)[82]	1.89(±0.05)
DQ15	1.65(±0.05)	1.64(±0.05)
DQ20	1.51(±0.05)	1.51(±0.05)

We find that DQ10, DQ15, and DQ20 have diffusion coefficients of ~1.88(±0.05)×10^{-10} m^2/s, ~1.65(±0.05)×10^{-10} m^2/s, and ~1.51(±0.05)×10^{-10} m^2/s respectively. These values are consistent with the diffusion of monomeric peptides with a similar number of amino acids, such as Aβ(1-28) (1.39×10^{-10} m^2/s).[16] From our diffusion coefficient measurements we roughly estimate that the Stokes radii of DQ10, DQ15, and DQ20 are ~10.5 Å, ~12.0 Å, and ~13.2 Å, respectively. These data indicate that DQ10, DQ15, and DQ20 are monomeric at 1 mg/mL in aqueous solution.

Our measured diffusion coefficients of DQ peptides monotonically decrease with peptide length. It is well known that that the reciprocal of the diffusion coefficient is linearly related to the van der Waals volume of the solute.[206–209] We find that this relationship also holds for DQ10-20 peptides (see Supporting Information). Thus, our observed decrease in diffusion coefficient for longer DQ peptides results from an increase in the monomeric peptide van der Waals volume.

DOSY measurements of NDQ10 indicate diffusion coefficients of $1.94(\pm0.05)\times10^{-10}$ m^2/s for Gln resonances and $1.42(\pm0.05)\times10^{-10}$ m^2/s for Lys resonances. The measured diffusion coefficient of the NDQ10 Gln residues is the same as that of the monomeric peptide DQ10 (within error). In contrast, the Lys residues of NDQ10 have a slower observed diffusion coefficient compared to DQ10. This observation is discussed in detail below.

DOSY measurements of NDQ15 report diffusion coefficients of $1.45(\pm0.05)\times10^{-10}$ m^2/s for Gln resonances and $0.75(\pm0.05)\times10^{-10}$ m^2/s for Lys resonances. The NDQ15 Gln and Lys diffusion coefficients are significantly less than that of the monomeric DQ15 peptide. In addition, the ^1H NMR spectrum of NDQ15 shows a broad background that is not observed for the other peptides studied (see Figure E.1). This background likely results from peak broadening of oligomeric species that results from long rotational correlation times.

To test the hypothesis that the broad background in the NDQ15 NMR spectrum results from oligomeric species, we collected ^1H and DOSY NMR spectra of NDQ15 before and after ultracentrifugation (355,524 X g, 30 min). We find that after centrifugation, the NMR background of the NDQ15 ^1H NMR spectrum is reduced, but not eliminated (see Figure E.1). This indicates that the NDQ15 NMR background results from the presence of oligomers. In addition, we find that after centrifugation the NDQ15 Gln residues have a diffusion coefficient of $1.64(\pm0.05)\times10^{-10}$ m^2/s and Lys residues have a diffusion coefficient of $1.25(\pm0.05)\times10^{-10}$ m^2/s. Thus, the observed NDQ15 diffusion coefficients after centrifugation are closer to that of the monomeric DQ15 peptide.

As mentioned above, the Lys diffusion coefficients of NDQ10 and NDQ15 are much smaller (~52% and ~73%, respectively) relative to that of the peptides' Gln residues. The dissimilarities between Gln and Lys diffusion coefficients can be explained by differences in the

mobilities of Gln and Lys residues in oligomers; numerous NMR experiments on Huntingtin exon 1 and small polyQ peptides show that flanking residues have significantly larger mobilities, compared to Gln residues, in polyQ fibrils.[53,54,157,210,211]

For example, Schneider et al. showed that Lys residues in $GK_2Q_{54}K_2$ and $GK_2Q_{38}K_2$ fibrils have greater mobility compared to the Gln residues.[210] Because, the Gln residues have low mobility and are incorporated into large fibril aggregates with a slow tumbling rate, the Gln signal is anisotropically broadened into the spectral background and is not observed. In contrast, the increased mobility of the Lys residues allows sufficient molecular tumbling to suppress anisotropic broadening even in large molecular weight fibrils. As a result, Schneider et al.[210] only observed Lys resonances in ^{1}H-^{13}C Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) NMR experiments, while chemical shifts of the Gln residues in the rigid fibril core are not observed because they are broadened into the spectral background.

The lower mobilities of the Gln residues in oligomers results in a Gln ^{1}H signal biased toward monomeric and low molecular weight oligomeric species. As a result, the measured Gln diffusion coefficients appear similar to that of the monomer. In contrast, the increased mobilities of the Lys residues in oligomeric species results in a Lys ^{1}H signal with contributions from monomer, low molecular weight oligomers, and high molecular weight oligomers. This results in a measured Lys ^{1}H diffusion coefficient significantly lower than that of the Gln residues.

This explanation is confirmed by the fact that after centrifugation of NDQ15 the Lys diffusion coefficient increases significantly more than the Gln diffusion coefficient. This result shows that the Lys resonances contain a much larger contribution from oligomeric species compared to Gln resonances. Because the Lys diffusion coefficients of both NDQ10 and NDQ15 are smaller than that of their respective monomers (DQ10 and DQ15), we conclude that both

NDQ15 and NDQ10 samples contain oligomers. In contrast, DQ10, DQ15, and DQ20 all show essentially the same diffusion coefficient for their Lys and Gln residues, confirming that these peptides are monomeric.

As discussed above, the Gln ^{1}H NMR peaks of NDQ10 and NDQ15 predominantly originate from the monomeric species because the ^{1}H resonances of oligomeric species are broadened into the spectral background. In contrast, the lysine residues have relatively high mobility even in oligomeric conformations and are less affected by anisotropic broadening. Assuming that the Gln ^{1}H NMR peak area originates from only the monomer species and that the Lys ^{1}H NMR peak area includes resonance from all monomeric and oligomeric species, we can roughly estimate the monomer concentration in NDQ15 and NDQ10. We normalized the Gln 2.31 ppm ^{1}H peak area to the area of the Lys 1.36 ppm ^{1}H peak for DQ10, DQ15, NDQ10, and NDQ15. We then compared the normalized Gln 2.31 ppm peak area of NDQ10 and NDQ15 to that of the completely monomeric DQ10 and DQ15 peptides. Using the procedure described above, we estimate that NDQ15 is ~59% monomeric while NDQ10 is ~80% monomeric.

When estimating the monomeric fraction of NDQ10 and NDQ15 we assume that the Gln H^{1} resonances only contain contributions from the monomer species. However, because the Gln residue diffusion coefficient is different for NDQ15 before and after centrifugation we know that the Gln H^{1} resonances must have an oligomer component. Therefore our reported monomer fractions for both NDQ10 and NDQ15 are overestimates.

To ensure that aggregation and fibrillization was not occurring during our DOSY measurements we collected ^{1}H NMR spectra of NDQ15 every ~30 min for 5 hrs. We monitored changes in the NDQ15 monomer concentration with time using the methods described above. We find that the monomer concentration does not change within the ~5 hour timescale of a

DOSY measurement (Figure E.5). This result confirms that our DOSY measurements are collected on a stable system.

It is important to note that our observation that NDQ10 contains ~80% monomer fractions differs from the conclusions of Punihaole et al.[82] who previously concluded that NDQ10 was 100% monomeric. This discrepancy results because Punihaole et al. previously examined the diffusion coefficients of only the Gln residues of NDQ10, which have a diffusion coefficient characteristic of monomeric peptide. Therefore, they concluded that NDQ10 was monomeric.

9.2.1.1 Structure of Peptide in PolyQ Oligomers

Previously, Jakubek et al.[193,204] and Punihaole et al. [82] quantitatively examined the secondary structures of NDQ10, NDQ15, and NDQ20 using UVRR and CD. They showed that NDQ10, NDQ15, and NDQ20 all have the same β-strand-like secondary structure with essentially identical UVRR and CD spectra. As discussed above, we find that the non-fibrillar oligomer population increases with peptide length for solution-state NDQ10-20 with NDQ10, NDQ15, and NDQ20 containing roughly 80%, 59%, and 0% monomeric fraction respectively. Because no changes in secondary structure were observed with increased oligomer population, we conclude that the monomeric peptide has the same secondary structure as that in oligomers. Thus, the presence of β-strand structure is present in polyQ monomer and non-fibrillar oligomers. Also, this result confirms that the β-strand-like structure exists prior to the formation of fibrils. This result is in agreement with that of Nagai et al.[16] who showed that the polyQ tract in thio-polyQ fusion proteins form soluble β-sheet structures before the formation of amyloid-like fibrils.

9.2.2 UV Resonance Raman

We used UVRR spectroscopy to investigate the peptide-peptide interactions of the disaggregated and non-disaggregated forms of Q10, Q15, and Q20. Recently, UVRR has become a powerful tool for investigating protein structure and solvation.[62,63] We use ~204 nm excitation that is in resonance with the $\pi \rightarrow \pi^*$ absorption band of amide groups, such as the primary amide Gln side chains and the secondary amide backbone peptide bonds. This allows for selective Raman enhancement of vibrations that couple to the resonant electronic transition. This includes the amide I (AmI) vibration of the primary amide Gln side chain (AmI^P) and the secondary amide peptide backbone (AmI^S). The AmI^S and AmI^P bands consist predominately of C=O stretching motions and have frequencies that are sensitive to solvation and hydrogen bonding environments of the amide carbonyl groups.[72,81,89]

Hydrogen bonding interactions with an amide carbonyl group decreases the C=O bond force constant causing an AmI frequency downshift.[81,89] In addition, changes in the carbonyl group's dielectric environment affects the frequency of the AmI bands. Solvents with a higher dielectric constant increase the stability of the amide dipolar resonance structure resulting in a weaker C=O bond force constant and a downshift in the AmI frequency.[72,89] Thus, the AmI frequencies of polyQ provide insight into the dielectric environment and hydrogen bonding interactions of the peptide backbone and Gln side chain amide groups.

Previously, Wang et al.[93] showed that the interaction enthalpy (ΔH_{int}) between a backbone amide carbonyl group and its environment is linearly correlated to the AmI^S frequency. Similarly, Punihaole et al.[81] showed that the Gln side chain ΔH_{int} is linearly correlated to the

AmI^P frequency. Using these correlations the ΔH_{int} for backbone and Gln side chain carbonyl groups can be estimated from experimentally obtained AmI^S and AmI^P bands, respectively.[81] Because the ΔH_{int} is dominated by strong hydrogen bonding interactions in protic environments, the ΔH_{int} values are an estimate of the hydrogen bonding strength.[81,152]

Estimating ΔH_{int} from AmI^P and AmI^S frequencies requires the assumption that transition dipole coupling (TDC) does not affect the AmI band frequencies. The backbone AmI^S band frequency is strongly dependent on TDC and changes depending on the peptide's secondary structure. Therefore determination of backbone amide ΔH_{int} is complicated. In contrast, the AmI^P frequencies of solution-state polyQ peptides show negligible dependence on TDC, and ΔH_{int} can be estimated for Gln side chains.[81] Punihaole et al.[81] previously derived the following equation to estimate the ΔH_{int} for Gln side chains:

$$\mathbf{AmI^P(cm^{-1}) = 1730(cm^{-1}) + (12(cm^{-1}\,kcal^{-1}\,mol))(\Delta H_{int}\,(kcal\,mol^{-1}))} \qquad \mathbf{9.1}$$

where AmI^P is the frequency of the AmI^P band and ΔH_{int} is the interaction enthalpy of the carbonyl group. Using this equation, Punihaole et al. previously estimated ΔH_{int} for a variety of hydrogen bonding interactions in polyQ peptides.[81]

For polyQ peptides, primary amide Gln carbonyl groups involved in side chain-water hydrogen bonding have an AmI^P frequency of ~1680 cm^{-1} and a corresponding ΔH_{int} of ~4.3 kcal/mol.[81] In contrast amide groups involved in either side chain-side chain or side chain-backbone hydrogen bonding have an AmI^P frequency of ~1660-1665 cm^{-1} and a corresponding ΔH_{int} of ~5.4-5.9 kcal/mol.[81] Similarly, the backbone AmI^S frequency of polyQ peptides is at

~1675 cm^{-1} when hydrogen bonded to water and is downshifted to ~1660-1665 cm^{-1} when involved in backbone-side chain or backbone-backbone hydrogen bonding.[46,193]

Because the AmI^P band frequency is similar for both side chain-side chain and side chain-backbone hydrogen bonding, we cannot differentiate between these two hydrogen bonding interactions. Also, the AmI^S band frequency for backbone-backbone and backbone-side chain hydrogen bonding is similar preventing their discrimination. Therefore, in this work, side chain-peptide hydrogen bonding indicates that the Gln side chain C=O groups are hydrogen bonded to peptide backbone NH groups and/or Gln NH_2 groups. Likewise, backbone-peptide hydrogen bonding indicates that the backbone C=O groups are hydrogen bonded to peptide backbone NH groups and/or Gln NH_2 groups.

The AmI^P and AmI^S bands of polyQ peptides overlap making it difficult to determine their individual frequencies. Xiong et al. showed that UVRR enhancement deeper in the UV (~197 nm) increases the resonance enhancement of the primary amide bands compared to that of the secondary amide vibrations.[57] As a result, for ~197 nm excitation, the Gln side chain bands are more enhanced relative to that at ~204 nm excitation. By subtracting the 204 nm UVRR spectrum from the 197 nm spectrum, we can highlight the UVRR bands of the Gln side chain. Also, by subtracting a 197-204 nm difference spectrum from the 204 nm spectrum we can highlight secondary amide vibrations.

9.2.2.1 Side Chain Hydrogen Bonding Interactions of PolyQ Peptides

Table 9.2 reports the measured AmI^P frequencies and ΔH_{int} (calculated using eq. 1) of the Q10, Q15, and Q20 solution-state polyQ peptides. Punihaole et al.[82] previously examined the AmI^P bands of DQ10 and NDQ10 while Jakubek et al.[193] previously examined the AmI^P bands

of DQ20 and NDQ20. Here we investigate the AmI^P bands of NDQ15 and DQ15 to elucidate the repeat length dependence of side chain polyQ hydrogen bonding interactions. The 197-204 nm UVRR difference spectra are shown in Figure 9.2.

Table 9.2: AmI^P Frequencies of NDQ and DQ Peptides Obtained from 197-204 nm UVRR Difference Spectra.

	AmI^P (cm^{-1})	ΔH_{int} (kcal/mol)
$NDQ10^{82}$	~1660, ~1679*	-5.8, -4.3*
NDQ15	~1661, ~1682*	-5.8, -4.0*
$NDQ20^{193}$	~1662	-5.6
$DQ10^{82}$	~1680	-4.2
DQ15	~1679	-4.3
$DQ20^{193}$	~1681	-4.1

*, Multiple values indicate multiple AmI^P bands and their respective ΔH_{int} values

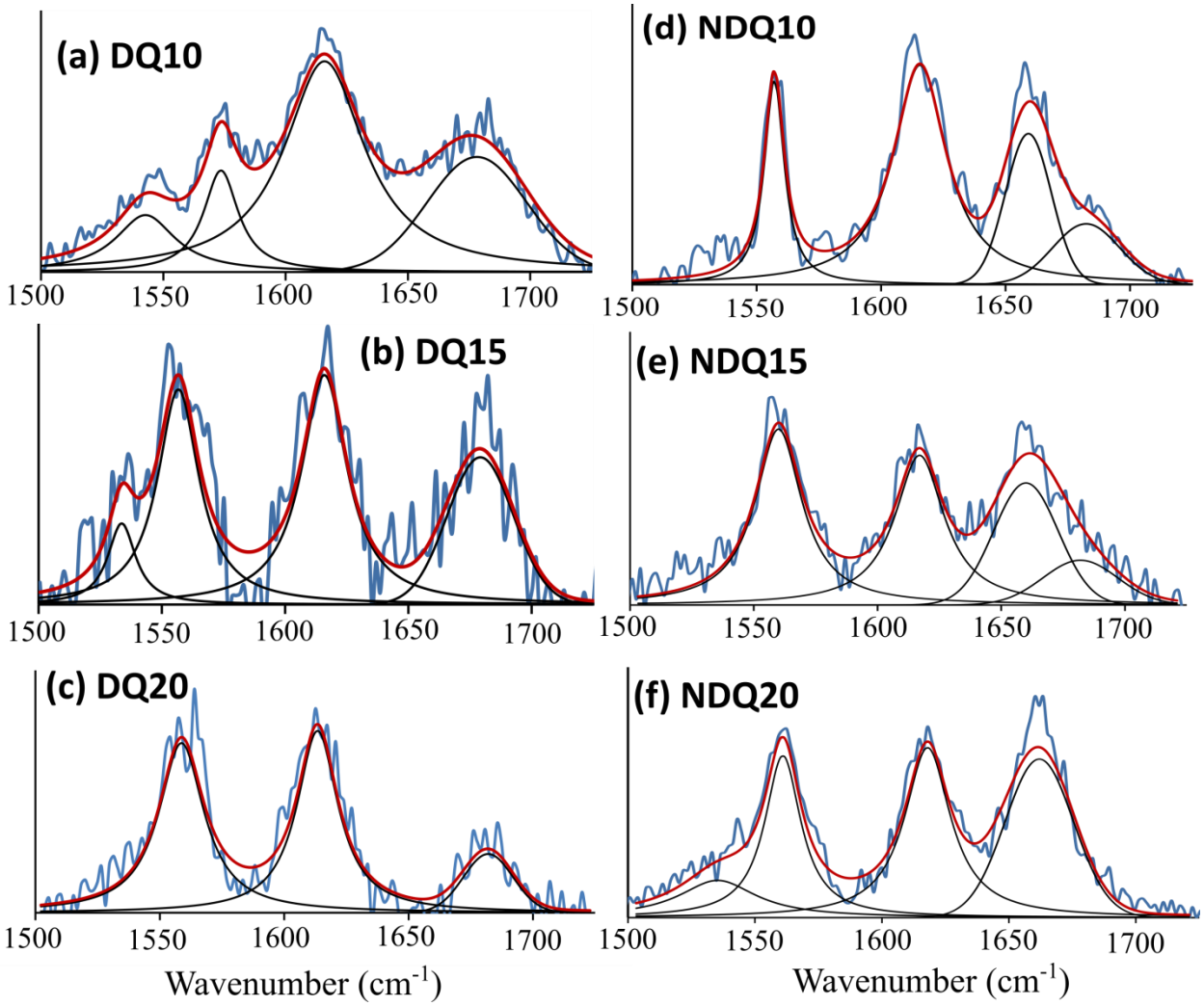

Figure 9.2: 197-204 nm difference spectra of (a) DQ10, (b) DQ15, (c) DQ20, (d) NDQ10, (e) NDQ15, and (f) NDQ20. The AmI^P band is found between 1650 and 1700 cm^{-1}. The Peak at ~1615 cm^{-1} is the $AmII^P$ band. The feature at ~1560 cm^{-1} is due to subtraction of the overlapping $AmII^S$ band (~1550 cm^{-1}) and O_2 stretching band (~1560 cm^{-1}) of atmospheric oxygen. The NDQ20 and DQ20 spectra were adapted with permission from Jakubek, R. et al. (2019). J. Phys. Chem. B, 123 (8), 1749-1763. Copyright 2019 American Chemical Society.

The AmI^P bands of DQ10[82], DQ15, and DQ20[193] are found at ~1680 cm^{-1}, which is indicative of Gln side chains hydrogen bonded to water.[81] Also, we do not observe an AmI^P band at ~1660 cm^{-1} for DQ10, DQ15, or DQ20 indicating that there is negligible side chain-peptide hydrogen bonding in these peptides. These results are consistent with our DOSY measurements showing that DQ10-20 peptides are monomeric and contain no oligomeric species, regardless of the polyQ tract length (see above). Using eq. 1 the ΔH_{int} of the side chain-water hydrogen bonds in DQ10, DQ15, and DQ20 are estimated to be ~-4.2 Kcal/mol[81], ~-4.3 Kcal/mol and ~-4.1 kcal/mol[193], respectively.

In contrast, NDQ10 and NDQ15 both contain AmI^P bands at ~1660 cm^{-1} and ~1680 cm^{-1}. As discussed above, the ~1660 cm^{-1} AmI^P band is characteristic of side chain-peptide hydrogen bonding while the ~1680 cm^{-1} AmI^P band is characteristic of side chain-water hydrogen bonding. This result indicates that NDQ10 and NDQ15 both contain a population of side chain-peptide and side chain-water hydrogen bonds. Using eq. 1, Punihaole et al. estimated that the ΔH_{int} for the side chain-water and side chain-peptide hydrogen bonds of NDQ10 are -4.3 kcal/mol and -5.8 kcal/mol, respectively. Using eq. 1 we estimate that the ΔH_{int} for the side chain-water and side chain-peptide hydrogen bonds of NDQ15 are ~-4.0 kcal/mol and ~-5.8 kcal/mol, respectively.

We find that the relative intensity of the 1680 cm^{-1} AmI^P band (compared to the 1660 cm^{-1} AmI^P band) is lower for NDQ15 compared to NDQ10. This indicates that NDQ15 contains fewer side chain-water hydrogen bonds and more side chain-peptide hydrogen bonds compared to NDQ10. This result is consistent with our DOSY data showing that NDQ15 contains a larger population of oligomers compared to NDQ10.

Previously, Jakubek et al.[193] examined the AmI^P band of NDQ20. NDQ20 is insoluble in water and Raman spectra were collected of the undissolved NDQ20 peptide in water. They found that NDQ20 contains an AmI^P band at ~1662 cm^{-1} with no observed band at ~1680 cm^{-1}. This indicates that NDQ20 contains side chain-peptide hydrogen bonding with negligible side chain-water hydrogen bonding. This observation is a continuation of the trend of longer NDQ peptides containing increased oligomerization and side chain-peptide hydrogen bonding.

9.2.2.2 Backbone Hydrogen Bonding Interactions of PolyQ Peptides

Table 9.3 reports the measured AmI^S frequencies of the Q10, Q15, and Q20 obtained from their 204-(197-204) nm UVRR difference spectra (Figure 9.3). Jakubek et al.[193] previously examined the AmI^S bands of DQ20 and NDQ20. Here we investigate the AmI^S bands of NDQ10, DQ10, NDQ15, and DQ15 to elucidate the repeat length dependence of backbone polyQ hydrogen bonding.

Table 9.3: AmI^S Frequencies of NDQ and DQ Peptides Obtained from 204-(197-204) nm Difference UVRR spectra.

	AmI^S (cm^{-1})
NDQ10	~1665
NDQ15	~1665
NDQ20	~1672[193]
DQ10	~1675
DQ15	~1675
DQ20	~1675[193]

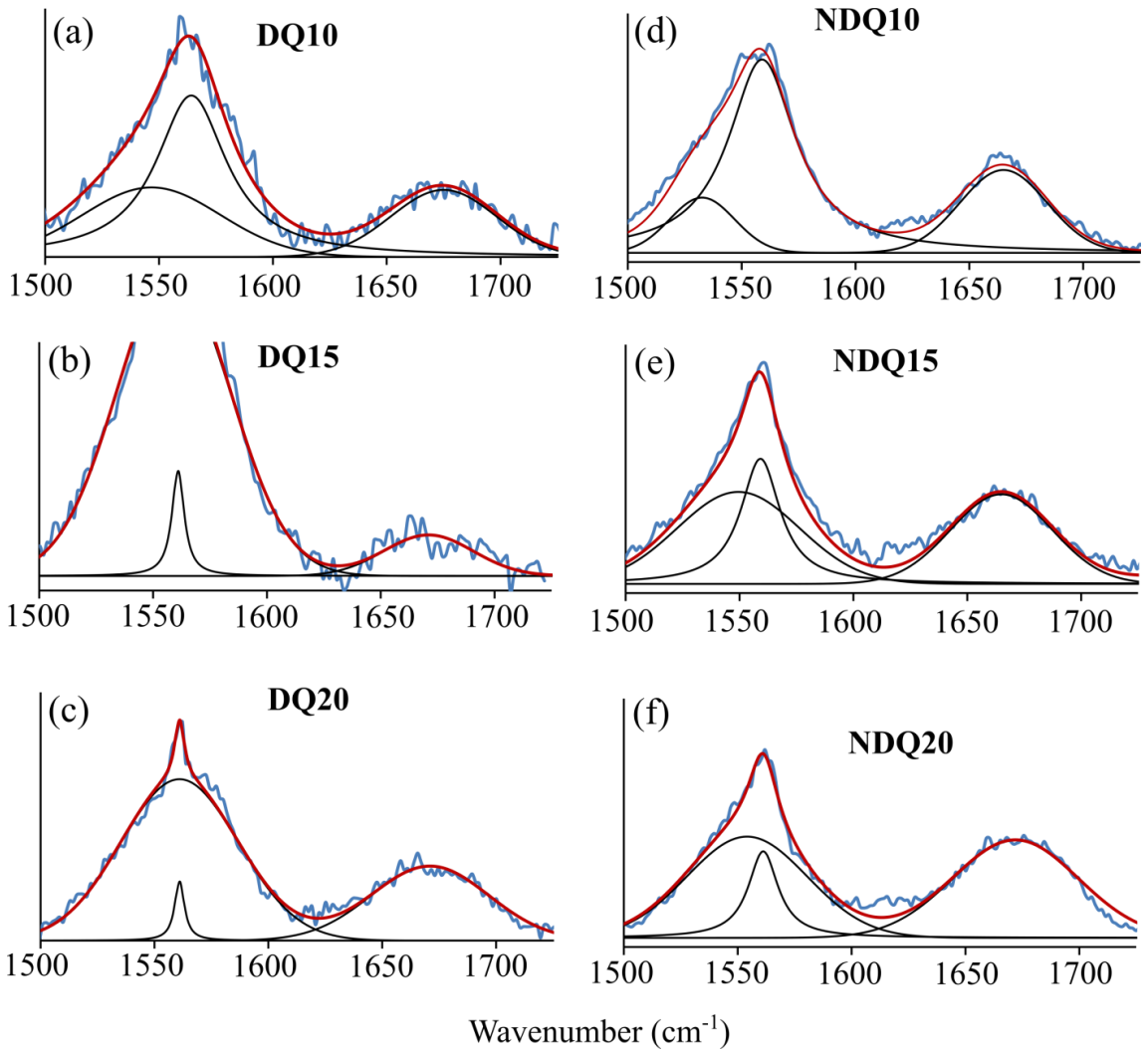

Figure 9.3: 204-(197-204) nm difference spectra of (a) DQ10, (b) DQ15, (c) DQ20, (d) NDQ10, (e) NDQ15, and (f) NDQ20. The AmI^{P} band is found between 1650 and 1700 cm^{-1}. The feature at ~1560 cm^{-1} is due to subtraction of the overlapping $AmII^{S}$ band (~1550 cm^{-1}) and O_2 stretching band (~1560 cm^{-1}) of atmospheric oxygen. The NDQ20 and DQ20 spectra were adapted with permission from Jakubek, R. et al. (2019). J. Phys. Chem. B, 123 (8), 1749-1763. Copyright 2019 American Chemical Society.

As discussed above, we previously showed that the AmI^S band of polyQ peptides is found at ~1675 cm^{-1} when the backbone amides are hydrogen bonded to water and at ~1660-1665 cm^{-1} when the backbone amides are hydrogen bonded to the peptide backbone or side chain groups.[81,193] From the 204-(197-204) nm UVRR difference spectra, we find that DQ10, DQ15, and DQ20 all have AmI^S band frequencies at ~1675 cm^{-1}. These frequencies indicate that DQ10, DQ15, and DQ20 have a peptide backbone involved in backbone-water hydrogen bonding.

Also, DQ10, DQ15, and DQ20 do not have AmI^S bands at ~1660-1665 cm^{-1} indicating that the peptides do not contain a significant population of backbone-backbone or backbone-side chain hydrogen bonds. This result is in agreement with our DOSY data indicating that these peptides are monomeric in solution.

In contrast, we find that NDQ10 and NDQ15 both have AmI^S bands at ~1665 cm^{-1}. This frequency indicates that NDQ10 and NDQ15 contain backbone-backbone and/or backbone-side chain hydrogen bonding. These results are in agreement with our DOSY data showing that NDQ10 and NDQ15 both contain oligomeric species that must have strong inter-peptide hydrogen bonding.

As discussed previously,[193] NDQ20 is insoluble in water and will therefore have negligible backbone-water hydrogen bonding. The AmI^S band of NDQ20 is found at ~1672 cm^{-1} which is significantly upshifted from that of NDQ10 and NDQ15.[193] This upshift indicates that the backbone amides of NDQ20 are involved in weaker backbone-peptide hydrogen bonding interactions compared to that of NDQ10 and NDQ15.[193] Figure 9.4 shows a depiction of the side chain and backbone hydrogen bonding interactions elucidated in this study.

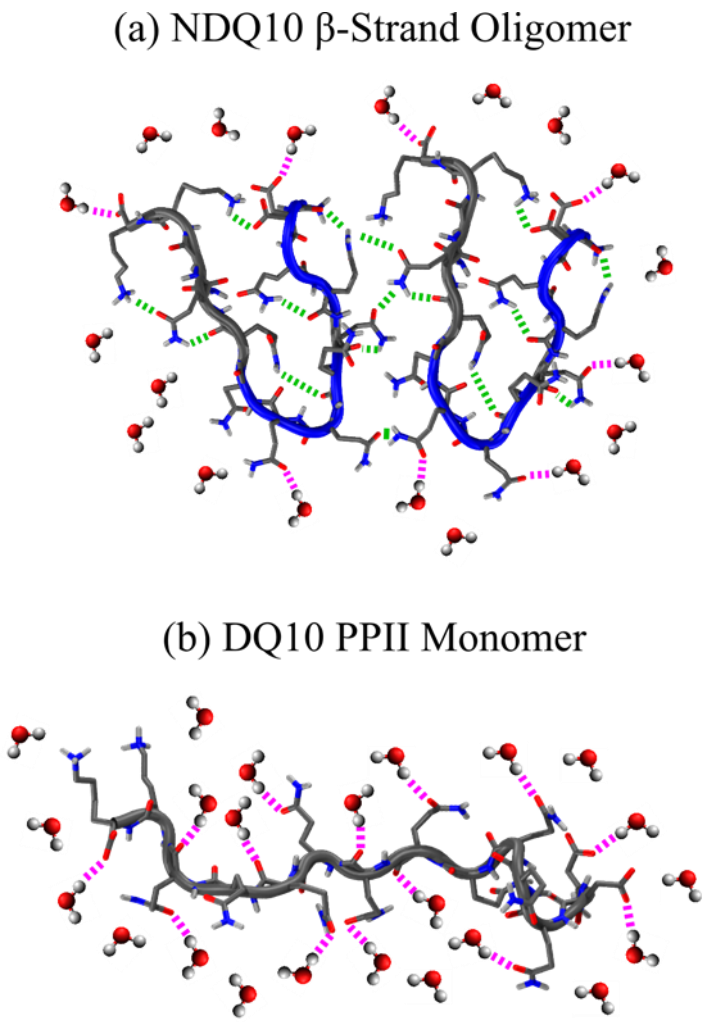

Figure 9.4: Depiction of the side chain and backbone carbonyl hydrogen bonding interactions in (a) NDQ oligomers and (b) DQ monomers. Green dotted lines indicate peptide-peptide hydrogen bonds and pink dotted lines indicate peptide-water hydrogen bonds.

9.2.3 Increased Oligomer Population is Correlated to Side Chain-Peptide Hydrogen Bonding

Above, we showed that, for NDQ peptides, the population of side chain-peptide hydrogen bonds and the population of non-fibrilllar oligomers increases with peptide length. In contrast, we do not observe an increase in backbone-peptide hydrogen bonding interactions for longer NDQ peptides with increased oligomer population. Therefore, we conclude that the formation of NDQ peptide oligomers is driven predominantly by side chain-peptide hydrogen bonding interactions.

Our conclusion is in agreement with previous work by Punihaole et al.[81] who showed that side chain-peptide hydrogen bonding interactions are stronger than backbone-peptide hydrogen bonding interactions in solution-state and fibril-state polyQ peptides. From this they concluded that side chain-peptide hydrogen bonding plays an important role in polyQ fibrillization and in stabilizing the fibril structure.

9.3 Comparisons to Other Studies

Our results are in agreement with several findings reported by Burra et al.[189] who examined the effects that disaggregation has on the structure and hydrogen bonding of polyQ peptides. They found that the disaggregation protocol disrupts the hydrogen bonding network present in NDQ peptides resulting in a monomeric DQ peptide. In agreement, we find that NDQ peptides have extensive peptide-peptide hydrogen bonding and solution-state oligomers while DQ peptides contain negligible peptide-peptide hydrogen bonding and are monomeric.

Also, Burra et al.[189] investigated the structure of DQ and NDQ polyQ peptides. They found that insoluble NDQ peptides exist in aggregates with β-sheet-rich structures that are less ordered compared to fibrils. This result is in great agreement with our results showing that NDQ10-15 monomers and oligomers as well as insoluble NDQ20[193] are in a β-strand-like conformation that is less ordered compared to β-sheet fibrils. A main discrepancy between our work and that of Burra et al.[189] is that they identify the DQ peptide structure as a random coil monomer while we find that DQ peptides are in a PPII-like conformation. This discrepancy results from the inability of the CD[198] and NMR techniques[202,203] used by Burra et al.[189] to identify the PPII conformation as discussed in the introduction.

9.4 Conclusions

We used DOSY to measure the diffusion coefficients and oligomerization of NDQ10, NDQ15, DQ10, DQ15, and DQ20 peptides. We find that DQ10, DQ15, and DQ20 all have diffusion coefficients consistent with monomeric peptides. In contrast, the diffusion coefficients of NDQ10 and NDQ15 indicate that both species form oligomers in water. From the DOSY spectra we estimate that ~80% of NDQ10 and ~59% of NDQ15 are in the monomer form. NDQ20 is insoluble in water, indicating increased oligomerization compared to NDQ15. Previously we showed, using UVRR and CD, that the secondary structures of NDQ10, NDQ15, and NDQ20 are all the same, indicating that both monomers and oligomers have the same β-strand-like secondary structure.[82,204]

We also used UVRR to investigate the hydrogen bonding interactions of NDQ10, NDQ15, NDQ20, DQ10, DQ15, and DQ20 peptides. We find that DQ10, DQ15, and DQ20 all

contain significant side chain-water and backbone-water hydrogen bonding with negligible peptide-peptide hydrogen bonding. In contrast we find that NDQ10, NDQ15, and NDQ20 contain a significant population of side chain-peptide and backbone-peptide hydrogen bonding. The extent of side chain-peptide hydrogen bonding increases with polyQ length and oligomer concentration for NDQ10-20. We do not observe a correlation between NDQ10-20 peptide length and backbone-peptide hydrogen bonding. Therefore, we conclude that the increased oligomerization of NDQ10-20 with polyQ peptide length results predominantly from an increase in side chain-peptide hydrogen bonding.

9.5 Experimental Methods

9.5.1 Materials

$D_2Q_{15}K_2$ (Q15) and $D_2Q_{20}K_2$ (Q20) were purchased form Thermo Scientific at ≥ 95% purity. TFA (purity ≥ 99.5%) and HFIP (99% pure) were purchased form Thermo Fisher Scientific and Acros Organics, respectively. D_2O was purchased from Cambridge Isotope Laboratories, Inc. with a deuterium atom purity of 99.9%.

9.5.2 Sample Preparation

NDQ peptides were prepared by suspending Q15 and Q20 peptides in water. Sterile Eppendorf Biopur® centrifuge tubes were used for NDQ sample preparation to prevent

impurities or dust particles from seeding aggregation. DQ peptides were prepared using the procedures developed by Chen et al.[31] and described by Punihaole et al.[82]

9.5.3 UVRR Instrumentation

The UVRR instrumentation used in this study was previously described in detail by Bykov et al.[103] We generated 197 nm laser excitation using the 4th harmonic of a tunable Ti:Sapphire laser (Positive Light Co.). ~204 nm light was generated by Raman shifting the third harmonic of a Nd:YAG Infinity laser (Coherent Inc.) in ~30 psi hydrogen. Raman scattering was dispersed using a home-built subtractive double monochromator and detected using a Spec10:400B CCD camera (Princeton Instruments) with a Lumogen E coating. Samples were spun in a Suprasil quartz NMR tube. A ~160° backscattering geometry was used to collect the Raman scattered light. Raman spectra where collected at room temperature (~20°C).

9.5.4 Diffusion Coefficient Measurements

All self-diffusion coefficient measurements were performed using a Bruker Avance III 600 MHz spectrometer with a BBFO probe with a Z-axis gradient. Temperatures were controlled to ±0.1 K accuracy using a Bruker BVT3000 temperature control system. Temperatures were calibrated with 4% Methanol in Methanol-d4 using the calctemp function in Bruker Topspin software. The pulsed field gradients were calibrated using a Bruker Doped Water sample to obtain a diffusion coefficient of 1.91×10^{-9} m^2/s at 25ºC.

The samples were thermally equilibrated at a set temperature of 25ºC for 30 min before the measurement. Measurements of the self-diffusion coefficients were performed by observing

^{1}H NMR signals. Self-diffusion coefficients were determined using a stimulated echo pulsed field gradient (STE-PFG) pulse sequence with bipolar gradients.[212,213] To obtain the self-diffusion coefficients, the peak intensity vs. gradient strength data were fit to the Stejskal−Tanner equation[105,214]:

$$\frac{\mathbf{I}}{\mathbf{I_0}} = \mathbf{e}^{-\boldsymbol{\gamma}^2\mathbf{g}^2\boldsymbol{\delta}^2\left(\Delta-\frac{\boldsymbol{\delta}}{3}\right)\mathbf{D}} \qquad \mathbf{9.2}$$

where, I and I_0 are the signal intensities with and without gradients, respectively, γ is the gyromagnetic ratio, g is the gradient strength, δ is the length of the gradient pulses, Δ is the diffusion time between the gradient pulses, and D is the self-diffusion coefficient. The gradient strength was varied between 0 and 50 G/cm while the duration of the gradient δ was held constant throughout the experiment. δ and Δ were set to 2.7 ms and 150 ms, respectively.

9.6 Acknowledgements

Funding for this work was provided by the University of Pittsburgh and the Defense Threat Reduction Agency (DTRA) HDTRA-09-14-FRCWMD (R.S.J., S.E.W., S.A.A.). S.E.W. gratefully acknowledges support through the Molecular Biophysics and Structural Biology NIH Training Grant (T32 GM 088119).

9.7 Supporting Information

Supporting information can be found in Appendix E and includes the following: DOSY NMR spectra of NDQ15, DQ15 and DQ20 with chemical shift assignments, DOSY signal decays for NDQ10 and DQ10, linear relationship between DQ10-20 peptide diffusion coefficients and peptide volume, and time dependence of the NDQ15 ^{1}H NMR spectra.

10.0 Conclusions

In this work we developed and utilized UVRR spectroscopy, as well as a variety of other techniques, to examine the structures of polyglutamine peptides in the solution-state and aggregate non-fibrillar state. In this section we highlight important findings that were discussed in detail in chapters 5.0 - 9.0.

Chapter 6.0 discussed the development of a technique allowing for the quantitative determination of glutamine side chain hydrogen bonding strengths from the frequency of the AmI^P band. We show that the AmI^P frequency is linearly correlated to the interaction enthalpy between a hydrogen bond donor and the Gln side chain C=O group hydrogen bond acceptor. Using this correlation, as well as work by Wang et al.,[93] we experimentally estimated the hydrogen bonding strengths for side chain-peptide, side chain-water, backbone-peptide, and backbone-water hydrogen bonding interactions in various polyQ structures. Our results were quantitatively compared to those of MD simulations and showed excellent agreement. We found that side chain-peptide hydrogen bonding is significantly stronger than backbone-peptide and peptide-water hydrogen bonding. These results indicate that side chain-peptide hydrogen bonding plays a crucial role in polyQ fibrillization and stabilizing the fibril structure. This work was the first, to our knowledge, to experimentally quantify the enthalpic favorability of polyQ hydrogen bonding interactions that are thought to play an important role in thermodynamically driving polyQ-rich proteins toward fibrils structures that are associated with neurodegenerative diseases.

In chapters 5.0 and 7.0 we utilized UVRR as well as other techniques to investigate the solution-state structures of Q10, Q15, and Q20. We found that the polyQ peptides can be

prepared in two different solution-state conformations. The non-disaggregated (NDQ) conformation is prepared by simply dissolving the peptide, produced via solid-phase synthesis, in water. This procedure results in a collapsed β-strand-like structure that contains significant peptide-peptide hydrogen bonding and is prone to aggregation. In contrast, the disaggregated peptide (DQ) is prepared following the procedure of Chen et al.[31] This preparation resulted in a PPII-like solution-state structure with negligible peptide-peptide hydrogen bonding and a high resistance to aggregation. We quantitatively compared Ψ angle distributions between UVRR measurements and polyQ structures obtained from metadynamics simulations. The experimental and computational results are in excellent agreement, resulting in an experimentally validated computational model for the structures of polyQ peptides.

Using CD spectroscopy we examined the dependence of the solution-state polyQ structure on the Gln repeat length. We found that longer DQ peptides, while predominantly in a PPII-like conformation, contain a larger subpopulation of the aggregation-prone β-strand-like conformation. In agreement, our metadynamics simulations found that the β-strand-like conformation is increasingly energetically favorable for longer polyQ peptides. We propose that the increase in stability of the aggregation-prone β-strand-like conformation for longer polyQ tracts may contribute to their increased aggregation kinetics, which is correlated with disease severity. These results are important because they provide insight into a possible mechanism by which longer polyQ peptides have faster aggregation kinetics that are correlated with earlier disease age of onset.

In addition, we examined the temperature dependence of the polyQ CD spectrum. We found that increased temperatures result in a decrease in the β-strand-like conformation while the PPII-like conformation is relatively insensitive to temperature change. This result agreed with

our metadynamics energy landscapes showing that the energy well for the PPII-like conformation was narrow while that for the β-strand-like conformation was broad, allowing for significant structural variability. It was previously proposed that increased structural variability promoted aggregation and fibrillization by increasing the number of intermolecular interactions and the probability of forming β-sheet structures that promote fibrillization.[37,38] Therefore, it is possible that the β-strand-like conformation is prone to fibrillization and aggregation because of its increased structural variability compared to that of the PPII-like structure.

In chapter 8.0 we examined the structure of the insoluble peptide NDQ20. We found that NDQ20 has a β-strand-like secondary structure similar to that of the soluble NDQ10 and NDQ15 peptides. In addition, we found that NDQ20 contains significantly weaker backbone-peptide hydrogen bonding compared to that of polyQ fibrils. In contrast, NDQ20 has side chain-peptide hydrogen bonding strengths similar to that observed in fibrils. These results indicate that the NDQ20 non-fibrillar aggregates are held together predominantly by side chain-peptide hydrogen bonding.

In addition, the Ψ angle distribution of NDQ20 (Ψ=~138°) differs from that of polyQ fibrils (Ψ=~148°). Also, the NDQ20 species shows more structural disorder compared to polyQ fibrils. Interestingly, we found that, even though NDQ20 is insoluble in water, it can convert to fibrils when exposed to water. These results suggest that the polyQ fibrillization mechanism may include pathways were non-fibrillar polyQ aggregates convert to fibrils and provide important insights into the mechanism of polyQ fibrillization.

Finally, in chapter 9.0 we used DOSY NMR and UVRR spectroscopies to investigate the concentrations and hydrogen bonding interactions of non-fibrillar, solution-state polyQ oligomers and their dependence on polyQ repeat length. We found that the disaggregated

peptides DQ10, DQ15, and DQ20 all have diffusion coefficients consistent with monomeric peptide and negligible peptide-peptide hydrogen bonding. In contrast, we found that both NDQ10 and NDQ15 contain a population of oligomers in water. From the DOSY data we estimate that NDQ10 contains ~80% monomer while NDQ15 contains ~59% monomer. As discussed in chapter 7.0, NDQ20 is insoluble in water (~0% monomer). Thus, the monomer population decreases with increasing peptide length for NDQ10, NDQ15, and NDQ20. Because we do not observe a significant change in secondary structure between NDQ10, NDQ15, and NDQ20 (see chapter 7.0), we concluded that both monomer and oligomer species have the same β-strand-like conformation. In addition, from UVRR data we found that the number of side chain-peptide hydrogen bonds in NDQ10, NDQ15, and NDQ20 increases with peptide length and oligomer population. However, we did not observe a correlation between peptide length (oligomer population) and backbone-peptide hydrogen bonding. Thus, we concluded that the increased oligomerization of longer non-disaggregated polyQ peptides is predominantly caused by an increase in side chain-peptide hydrogen bonding. These results are the first to show that non-disaggregated polyQ monomers and non-fibrillar oligomers have the same structure, and that non-fibrillar oligomers are formed predominantly through side chain-peptide hydrogen bonding.

Overall, the work presented here provides numerous important insights into the structures and hydrogen bonding interactions experienced by polyQ tracts. These insights greatly improve our understanding of polyQ aggregation and fibrillization, which is important in understanding the pathology of CAG repeat neurodegenerative diseases.

11.0 Future Work

11.1 Examining the PolyQ Length Dependence of the Fibril Structure

In this work we utilized UVRR spectroscopy to examine the structure and hydrogen bonding interactions of solution-state polyQ peptides and their dependence on the length of the polyQ tract. Future work will include examining the hydrogen bonding and structures of polyQ fibrils of Q15 and Q20. We will compare these structures to those previously determined for $Q10^{46}$ to examine the possibility of differences in the polyQ fibril structure for peptides with different polyQ tract lengths. PolyQ fibrils have been proposed as the toxic species in polyQ diseases. This work could provide insight into the role that polyQ fibrils play in CAG repeat diseases.

11.2 Examining PolyQ Structure of Longer PolyQ Tracts

CAG neurodegenerative diseases only occur when the polyQ tract in a protein exceeds a threshold length. The disease threshold generally occurs for polyQ tracts >~40 Gln residues.[1] For example, the critical repeat length for Huntington's disease is 36 Gln residues.[1] Because of this, there is great interest in comparing the structures of polyQ tracts below and above the disease threshold. We have shown that UVRR, coupled with other techniques, can be used to obtain detailed structural information on these systems. In the future we will expand our examination of polyQ solution-state structure to peptides containing longer polyQ tracts that span the disease

threshold repeat length for many CAG repeat diseases. This work will allow for the identification of structural and hydrogen bonding changes in longer polyQ peptides that may have relevance to CAG repeat disease pathology.

11.3 Examination of PolyQ Tract Structure with Flanking Polyproline Tract and in Huntingtin Exon 1

It is well-known that the structure and aggregation mechanism of polyQ peptides are modified by flanking amino acid sequences.[48,199] In the huntingtin protein, the polyQ tract is flanked by a C-terminal polyproline (polyP) sequence as well as an N-terminal sequence. In future work we will use UVRR to investigate the effect of these flanking sequences on the structure of the polyQ tract. This work will begin by examining the structure of polyQ with a polyP flanking sequence. Once this system is understood we will examine the structure of polyQ in the full huntingtin exon 1 containing both the C-terminal polyP tract and the N-terminal flanking residues. This work will elucidate polyQ conformations that may be important in the pathology of Huntington's disease.

11.4 Examination of Structural Changes During PolyQ Fibrillization

The measurement time of a polyQ UVRR spectrum is seconds to minutes. This allows for the collection of UVRR spectra, that detail polyQ secondary structure and hydrogen bonding, on time scales shorter than that of polyQ fibrillization. Thus, using UVRR, one can obtain structural

information on the polyQ sequence throughout the time course of fibrillization. We will utilize UVRR to examine the structural evolution of polyQ peptides as they form fibrils. This work will aim to identify any intermediate structures during polyQ fibrillization and to compare the intermediate structures across polyQ peptides of different length. The results of this work have the potential to provide direct insights into the fibrillization mechanism(s) of polyQ tracts.

Appendix A

Supporting Information for Chapter 5.0

A.1 UVRR Spectral Processing and Fitting

UVRR Spectral Processing

All UVRR spectra were processed using home-written MATLAB scripts to remove cosmic ray contributions, average, and calibrate spectra. The spectral contributions of water and Suprasil quartz from the NMR tubes were subtracted using a methodology described by Punihaole et al.[46]

Global Fitting of UVRR Spectra

The UVRR spectra were least-squares fit as the sum of *N* Gaussians and Lorentzians:

$$S(\nu) = \sum_{i}^{N} \left[f_i H_i e^{-\left(\frac{\nu-\nu_i}{w_i}\right)^2 (4ln(2))} + (1-f_i)\frac{H_i}{4(\frac{\nu-\nu_i}{w_i})^2+1} \right] \quad \textbf{A.1}$$

where $f_i = 1$ if the i^{th} band is a Gaussian, or 0 if the i^{th} band is a Lorentzian. The parameters H_i, ν_i, and w_i are the heights, center frequencies, and widths, respectively, of the Gaussian or Lorentzian

bands. The H_i, v_i, and w_i parameters were determined using a global fitting procedure where each parameter was fit to three different spectral replicates. The results of fitting our data using this global analysis are shown in Figure A.1.

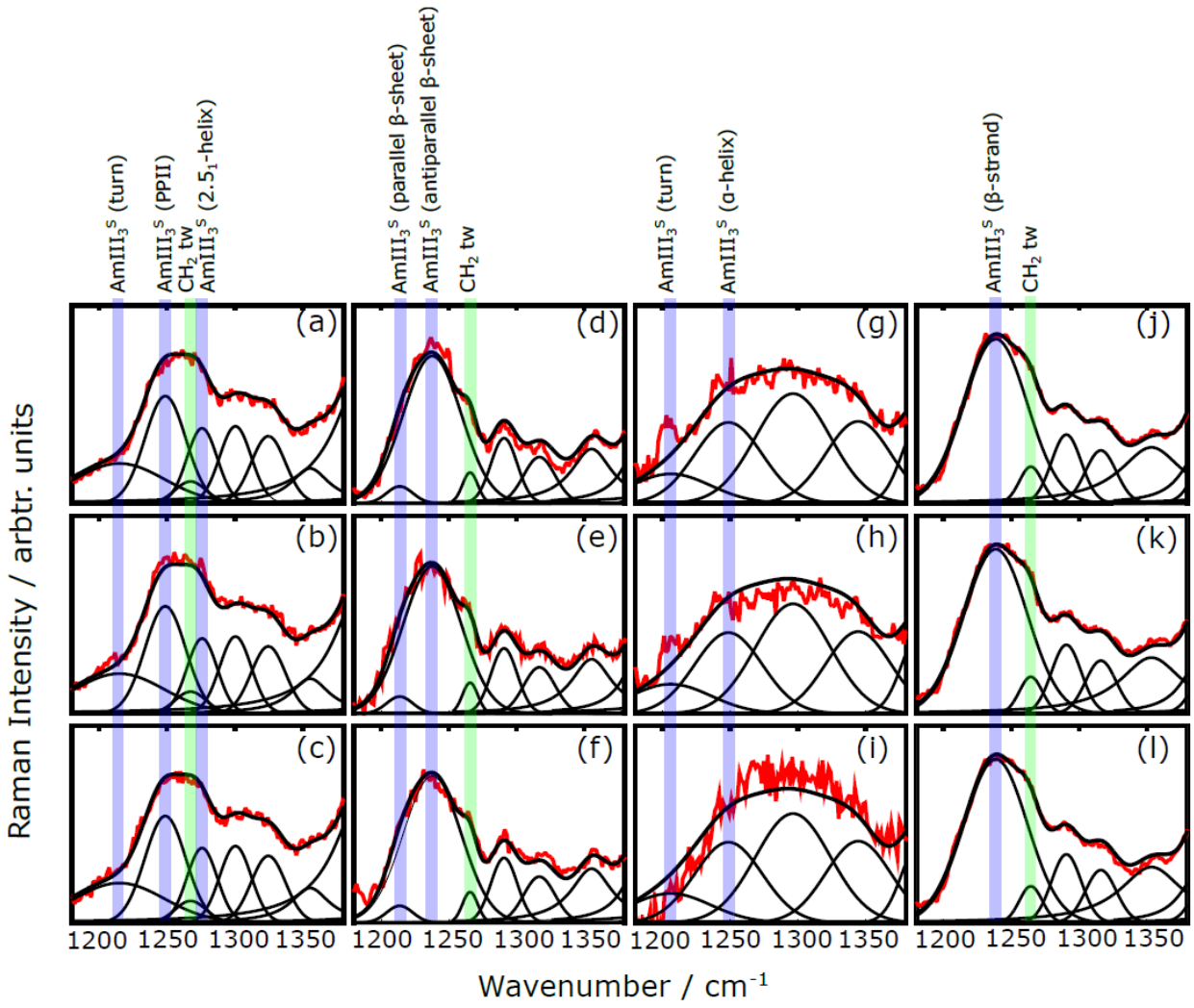

Figure A.1: Global fitting of UVRR spectral replicatesfor: (a–c) an ensemble of PPII-/2.5$_1$-helix-like conformations prepared from DQ10 peptides in pure water; (d–f) Predominately antiparallel β-sheet fibrils prepared from DQ10 peptides in acetonitrile/water mixtures where X_{CD3CN} = 0.74/X_{H2O} = 0.26; (g–i) α-helix-like structures prepared from DQ10 peptides in acetonitrile/water mixtures where X_{CD3CN} = 0.56/X_{H2O} = 0.44; and (j–l) β-strand-like structures prepared from NDQ10 peptides in pure water. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

Basis Spectra Fitting

The UVRR spectra shown in Figure 5.6a were modeled as a linear combination of individual basis spectra (shown in Figure 5.6b) representing PPII-like, α-helix-like, and β-sheet secondary structures. To do this, we utilized the classical multiple linear least-squares regression method described by Chi et al.:[215]

$$[S] = [R][S_B] \qquad \textbf{A.2}$$

where $[S]$ is the $(n \times 1)$ row vector that represents the intensities of the experimentally measured spectrum and $[S_B]$ is the $(n \times m)$ matrix composed of row vector elements that contain the different secondary structure basis spectra. The $(m \times 1)$ row vector, $[R]$, contains the least-squares scaling coefficients that represent the relative contributions of the $[S_B]$ basis spectra components to the Figure 5.6a spectra. The least-squares solution of $[R]$ for eq. A.2 is:[215]

$$[R] = \{[S_B]^T[S_B]\}^{-1}[S_B]^T[S] \qquad \textbf{A.3}$$

where $\{[S_B]^T[S_B]\}^{-1}[S_B]^T$ is the pseudo-inverse matrix of $[S_B]$. The results of modeling the Figure 5.6a spectra are shown in Figure A.2.

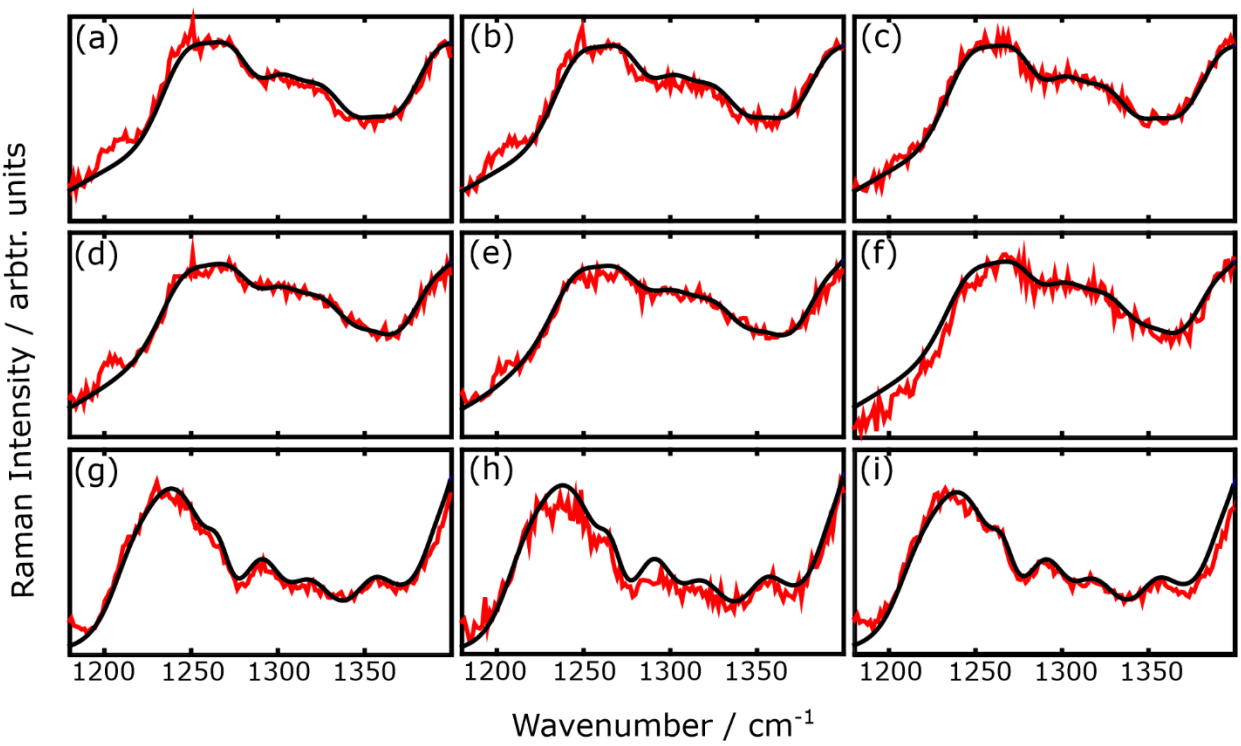

Figure A.2: Modeling of UVRR spectra of DQ10 in different acetonitrile/water mixtures as a linear combination of PPII-like, α-helix-like, and β-sheet basis spectra. (a–c) Modeling of 3 separate UVRR spectra of DQ10 peptides in water/acetonitrile mixtures where X_{H2O} = 0.76/X_{CD3CN} = 0.24; (d–f) Modeling of 3 separate UVRR spectra of DQ10 peptides in acetonitrile/water mixtures where X_{H2O} = 0.44/X_{CD3CN} = 0.56; and (g–i) Modeling of 3 separate UVRR spectra of DQ10 peptides in acetonitrile/water mixtures where X_{H2O} = 0.30/X_{CD3CN} = 0.70. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

A.2 **Determination of Ψ and χ_3 Angle Distributions**

The distributions of Ψ and χ_3 dihedral angles, shown in Figure 5.3 and Figure 5.7, were calculated using methodologies previously described in detail.[64,71] We assume that the inhomogenously broadened AmIII$^S_3$ and AmIII^P bands ($B(\nu)$) can be modeled as the sum of M Lorentzian bands with identical homogeneous linewidths:

$$B(\nu) = \frac{1}{\pi} \sum_{i}^{M} \frac{p_i \Gamma^2}{\Gamma^2 + (\nu - \nu_i)^2} \qquad \textbf{A.4}$$

where p_i is the probability for the i^{th} band to occur at center frequency ν_i and Γ is the homogeneous linewidth of the AmIII$^S_3$ or AmIII^P vibrations. We previously estimated from peptide crystals that Γ is ~7.5 cm^{-1} for the AmIII$^S_3$ and ~6.6 cm^{-1} AmIII^P.[71,74] After fitting the inhomogeneously broadened bands with Lorentzians, we calculate the different i^{th} frequencies of the AmIII$^S_3$ and AmIII^P band envelopes and their respective Ψ or χ_3 dihedral angle distributions.

A.3 **Correlating the AmIII$^S_3$ Frequencies to Ψ Angles**

We correlated the AmIII$^S_3$ PPII and 2.5$_1$-helix band frequencies of DQ10 to Ramachandran Ψ angles using eq. A.5:

$$\nu_i(\Psi,T) = 1256\ (\mathrm{cm}^{-1}) - 54\ (\mathrm{cm}^{-1})\sin(\Psi + 26\ ) - 0.11\ (\mathrm{cm}^{-1}/°\mathrm{C})(T - T_0) \qquad \textbf{A.5}$$

where T = 18°C is the sample temperature and T_0 = 0°C. As discussed in detail by Mikhonin et al.,[64] eq. A.5 is used for situations when the peptide bonds are fully exposed to water, such as in the case of PPII-like and 2.5_1-helices.

Eq. A.6 was utilized to calculate the Ψ Ramachandran angles from the AmIII^{S}_{3} band frequencies of NDQ10:

$$\nu_i(\Psi,T) = 1250\ (\mathrm{cm}^{-1}) - 54\ (\mathrm{cm}^{-1})\sin(\Psi + 26\ ) + 0.06\ (\mathrm{cm}^{-1}/°\mathrm{C})(T - T_0) \qquad \textbf{A.6}$$

where parameters T and T_0 are the same as in eq. A.5. Eq. A.6 is recommended by Mikhonin et al.[64] for situations when the hydrogen bonding states of the peptide bond N–H groups are unknown. This situation occurs, for example, in the case of turn-like structures, where the peptide bonds can engage in both inter-amide and amide-water hydrogen bonds.

A.4 **Determining the χ_3 Angles from the AmIII^{p} Frequencies**

The χ_3 angle distributions of DQ10 were determined previously by Punihaole et al.[71] For NDQ10, we fit the AmIII^{p} region of the 197 nm – 204 nm UVRR difference spectrum shown in Figure 5.4. Based on our previous studies of glutamine,[71] we assign the 1106 cm^{-1} band to the AmIII^{p} vibration. We used the following equation to calculate the NDQ10 χ_3 angle from its AmIII^{p} band frequency:

$$\nu_i(\chi_3) = 1076\ (\mathrm{cm}^{-1}) + 29\ (\mathrm{cm}^{-1})\cos(2\chi_3) + 9\ (\mathrm{cm}^{-1})\cos(\chi_3 + 99\ ) \qquad \textbf{A.7}$$

where ν_i is the ith AmIII^P frequency.

Equation A.7 was derived by Punihaole et al.[71] for use in situations when the hydrogen bonding and dielectric environments of the glutamine side chains are unknown. Our previous study on primary amide solvation showed that stronger (weaker) hydrogen bonding and higher (lower) dielectric environments upshift (downshift) the AmIII^P frequency.[72] Eq. A.7 effectively averages these two competing effects. We used eq. A.7 because NDQ10 exhibits significant inter-amide hydrogen bonding (of the C=O groups), as well water-amide hydrogen bonding (of the NH_2 groups).

Each AmIII^P frequency can be correlated to as many as four different χ_3 angles. As discussed in detail by Punihaole et al.,[71] χ_3 dihedral angles that are greater than +90° and less than -90° are nearly forbidden for glutamine side chains. Thus, for DQ10 and NDQ10, we only considered χ_3 angles between -90° and +90°.

A.5 **Hydrogen-Deuterium Exchange**

Hydrogen-Deuterium exchange (HX) measurements were used, in combination with DOSY NMR, to determine whether NDQ10 is monomeric or aggregated. Monomeric peptides are well-solvated and in intimate contact with water. Thus, HX exchange rates are relatively fast, ~10 min. In contrast, oligomeric or aggregated species exhibit peptide backbones that are shielded from solvent, and, thus, have significantly slower HX kinetics.[216] If there is only one

species present (either monomer or aggregate), then the HX kinetics will show a monoexponential time dependence. If multiple species are present, then the HX kinetics will be more complex.

HX measurements were carried out by initially preparing NDQ10 and DQ10 peptide solutions in H_2O, and then rapidly mixing in D_2O so that the final sample concentration was 0.3 $mg \cdot mL^{-1}$. The dead times for all mixing experiments was ~1 min. Sixty UVRR spectra were collected in succession with 1 minute accumulation time per spectrum. To increase signal-to-noise, while still maintaining acceptable temporal resolution, we averaged 3 consecutive spectra, resulting in kinetic time points every 3 min during the exchange. To further increase signal-to-noise, we smoothed the spectra using a Savitzky-Golay filter[217] with a second order polynomial over a 15 data point interval.

Deuteration of the peptide backbone decouples NH bending from C–N stretching in the AmII vibration. This decoupling results in the AmII' vibration, which consists mainly of C–N stretching. We determined the HX kinetics of DQ10 and NDQ10 in water by monitoring the AmII' band intensity at ~1470cm^{-1}. The AmII' band intensities were normalized to calculate the fraction of exchanged hydrogens (f_{HX}) using the following equation:

$$f_{HX} = \frac{I_{AmII'}(t) - I_{AmII'}(\infty)}{I_{AmII'}(0) - I_{AmII'}(\infty)} \qquad \textbf{A.8}$$

where $I_{AmII0}(t)$ is the intensity at 1470cm^{-1} at time t, $I_{AmII0}(0)$ is initial intensity at 1470cm^{-1} in the absence of HX, and $I_{AmII0}(\infty)$ is the intensity at 1470cm^{-1} at equilibrium.

Figure 5.5c shows f_{HX} as a function of time. To extract a time constant for HX, we fit the Figure 5.5c data to the following equation:

$$f_{HX}(t) = 1 - e^{-t/\tau} \quad \textbf{A.9}$$

where τ is the time constant of HX in min.

Our experiments show that HX kinetics of DQ10 and NDQ10 (Figure 5.5c) show time constants of 5.4(2) and 5.6(4) min, respectively. These time constants are consistent with those measured for other monomeric peptides.[121] Thus, we conclude from our HX experiments, as well as our DOSY NMR data and hydrogen bonding analysis, that both DQ10 and NDQ10 are monomers in solution.

A.6 **Circular Dichroism (CD) Measurements**

We measured the CD spectra of DQ10 peptide solutions in acetonitrile/water mixtures where X_{CH_3CN} = 0, 0.24, and 0.56, as shown in Figure A.3. We did not measure the CD spectra of DQ10 peptides in acetonitrile/water mixtures with higher mole fractions of acetonitrile due to the presence of visible aggregates. The CD spectra were measured using a Jasco J-710 spectropolarimeter and a quartz cuvette with a 0.2 mm pathlength. The CD spectra were collected every 0.2 nm interval and averaged over 5 scans. The spectra were smoothed using a Savitzky-Golay filter[217] with a second order polynomial over a 15 data point interval.

The CD spectrum of DQ10 in pure water shows a strong negative band at ~200 nm and a slightly negative band at ~220 nm. These features indicate that DQ10 adopts predominately PPII-like structures in water.[218,219] In agreement with the UVRR results, the CD spectral changes that occur between X_{CH_3CN} = 0 and X_{CH_3CN} = 0.56 are indicative of a PPII-like to α-helix

transition. As the solvent is enriched in acetonitrile, the ~220 nm band becomes more negative, while the ~200 nm band blue-shifts and becomes positive. At X_{CH_3CN} = 0.56, an additional negative band, characteristic of α-helix-like structures, appears at ~205 nm.

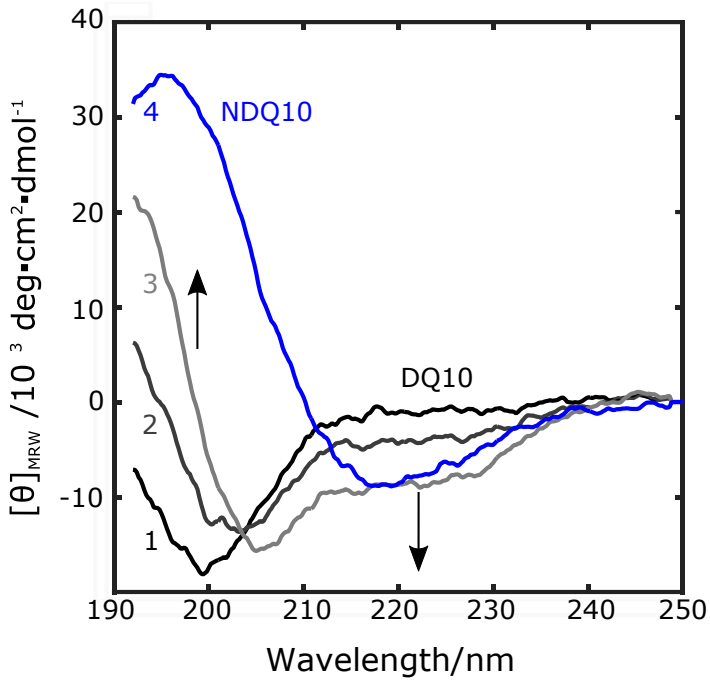

Figure A.3: CD spectra of DQ10 and NDQ10 peptides in different acetonitrile/water mixtures. (1) DQ10 in pure water. (2) DQ10 in an acetonitrile/water mixture, X_{CH3CN} = 0.24. (3) DQ10 in an acetonitrile/water mixture, X_{CH3CN} = 0.56. (4) NDQ10 in pure water. The spectra were smoothed using a Savitzky-Golay filter with a second order polynomial over a 15 data point interval. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

A.7 **Conformational Reversibility of NDQ10 in Acetonitrile Solutions**

We discovered that at high acetonitrile concentrations NDQ10 form low molecular weight β-sheet structures. To test the reversibility of these β-sheet structures upon rehydration, we dissolved NDQ10 in a 70% acetonitrile solution. This solution was then further diluted with either 70% acetonitrile or H_2O resulting in 0.3 $mg \cdot ml^{-1}$ NDQ10 solutions in 70% acetonitrile and in 7% acetonitrile respectively. UVRR (204 nm) spectra were collected for each solution. NDQ10 in 70% acetonitrile shows a UVRR spectrum characteristic of the β-sheet conformation of fibrils, while the 7% acetonitrile solution UVRR spectrum was characteristic of aqueous NDQ10 in its collapsed β-strand conformation. Thus, the β-sheet structure of NDQ10 at high acetonitrile concentrations is reversible upon rehydration.

A.8 **Supplementary Figures**

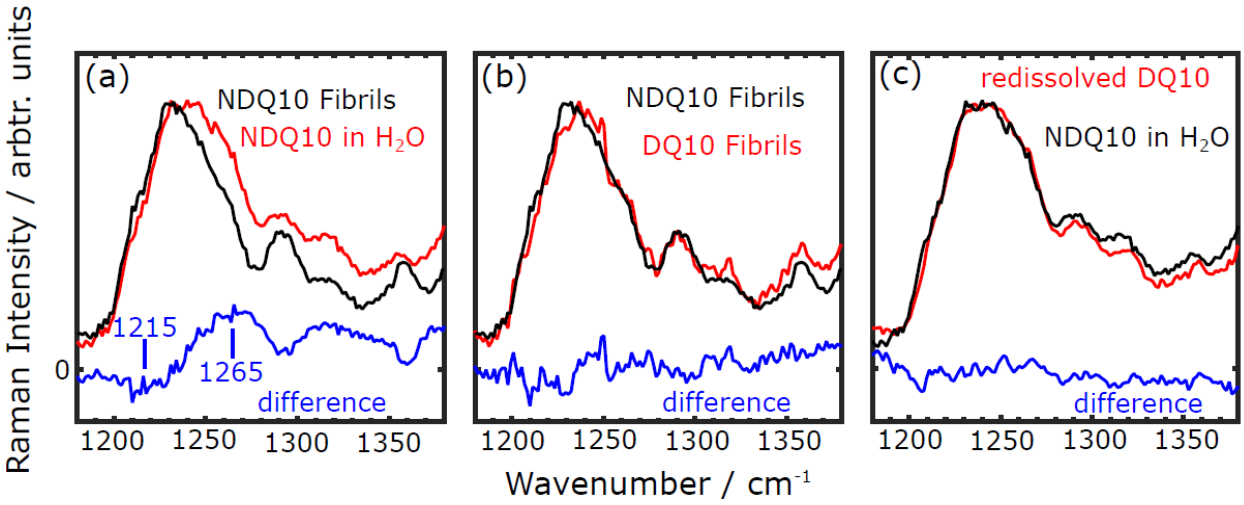

Figure A.4: UVRR Spectral comparisons of fibrils formed under different conditions. (a) Comparison of NDQ10 fibrils form in water (black) and monomeric NDQ10 in water (red). (b) Comparison of NDQ10 fibrils formed in water (black) and DQ10 fibrils (red) prepared in water. (c) Comparison of monomeric, aqueous NDQ10 (black) and DQ10 fibrils grown in acetonitrile and re-dissolved in water (red). Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

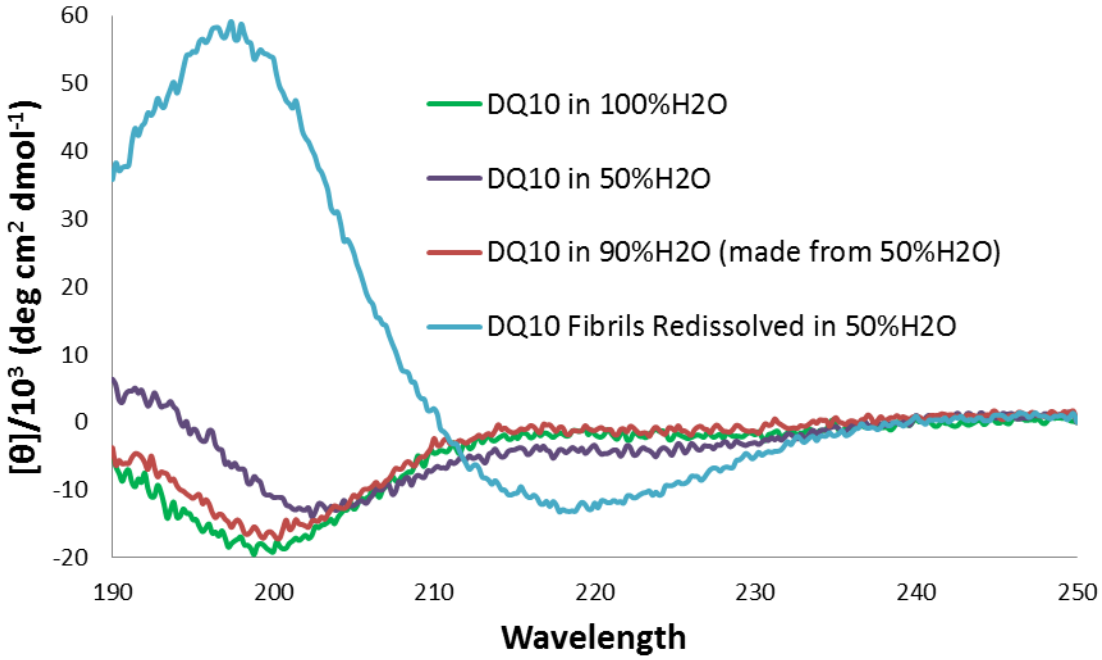

Figure A.5: CD spectra of DQ10 in different acetonitrile/water mixtures. The CD spectra are of DQ10 in (green) 100% water, (brown) 50% H_2O, (orange) 90% water after dilution from a 50% water solution, and (blue) 50% water of dissolved fibrils. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

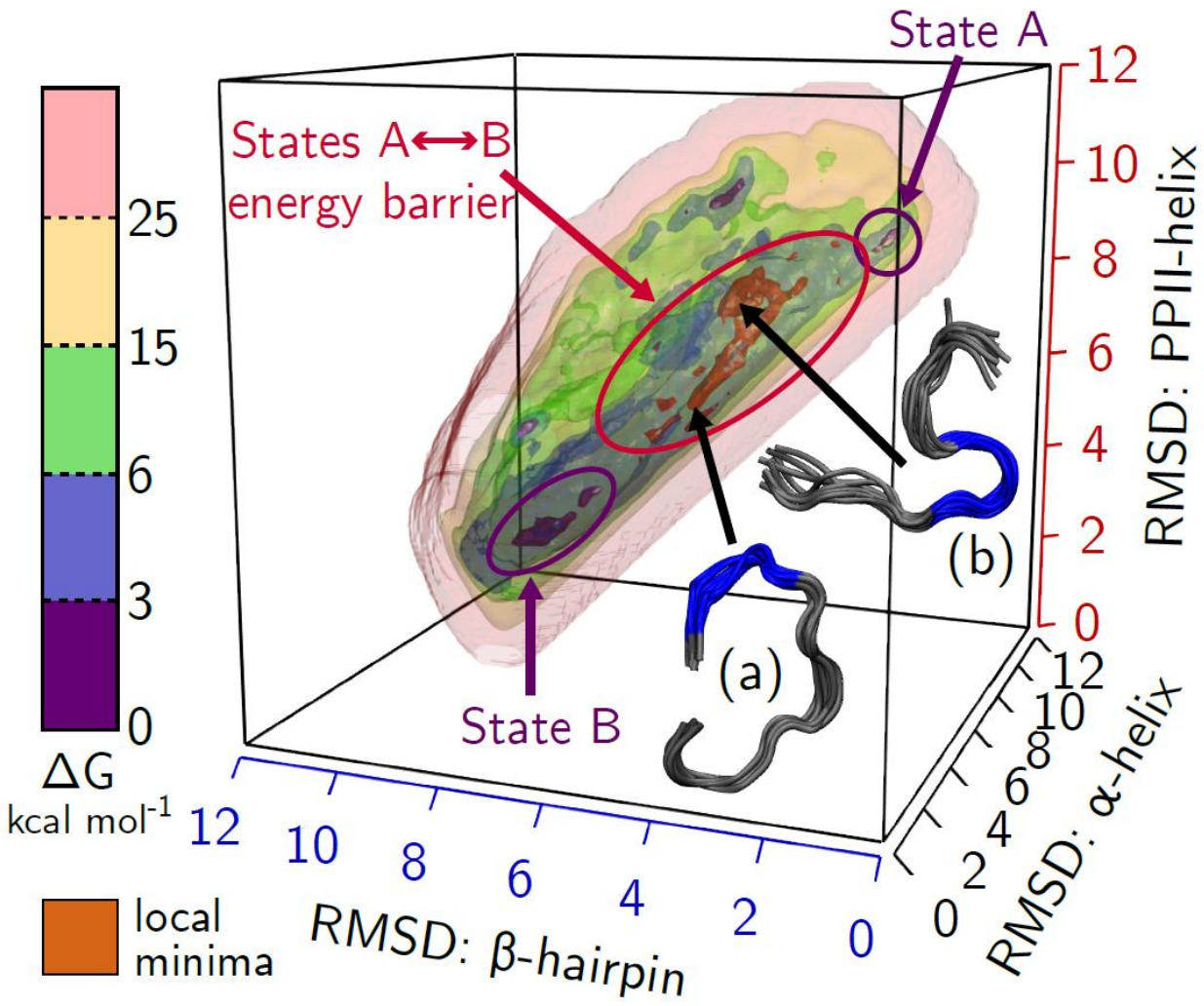

Figure A.6: Q10 conformers found along the free energy barrier separating states A and B. The energy barrier between states A and B is circled in red. Black arrows point to the location of representative transition structures on the landscape. Purple arrows show the positions of low energy states A and B on the landscape. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

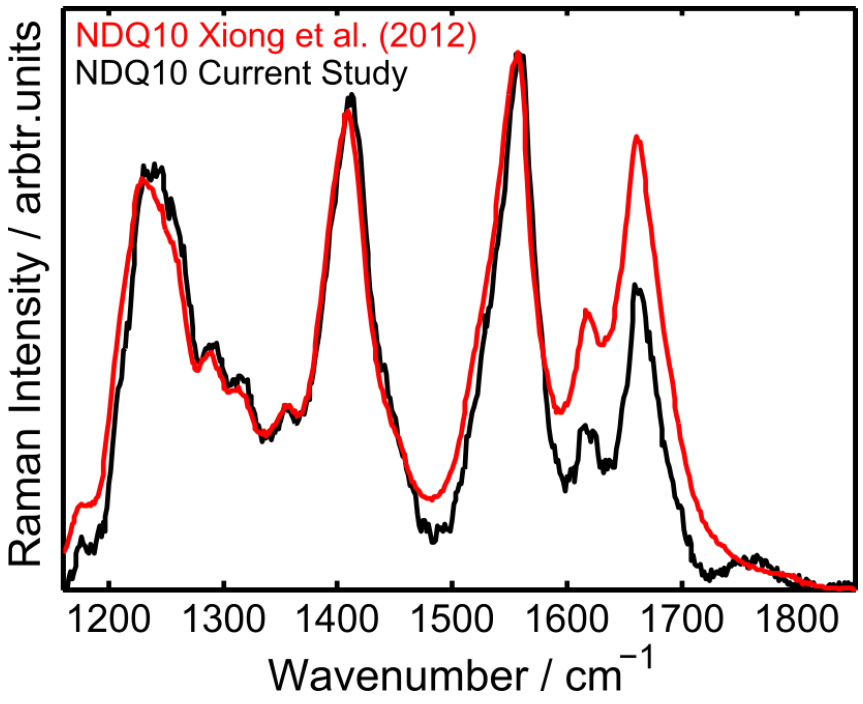

Figure A.7: Comparison of solution-state NDQ10 spectra previously reported by Xiong et al.[57] (red) and reported in this article (black). Spectra are normalized to the $AmIII_3^S$ band for comparison. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

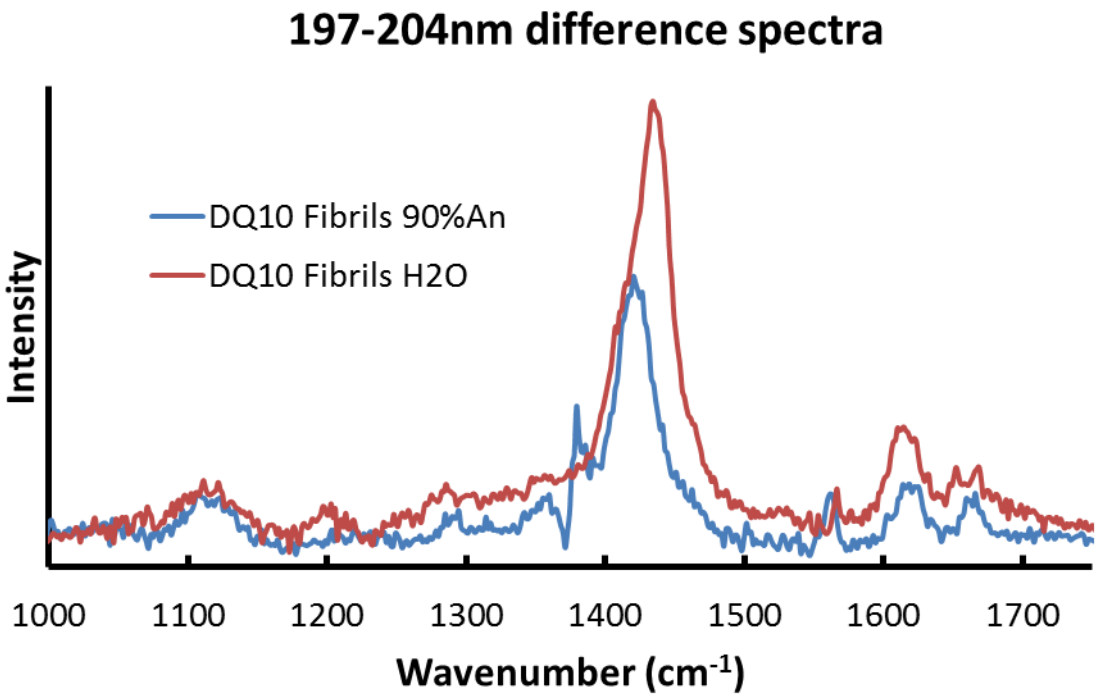

Figure A.8: Comparison of the 197-204 nm difference spectra of DQ10 fibrils prepared in (red) water[46] and (blue) 90% acetonitrile. The 197-204 nm difference spectrum highlights primary amide vibrations of the Gln side chain. The AmI^P band is at a similar frequency (~1664 cm^{-1}) for the two types of fibrils. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

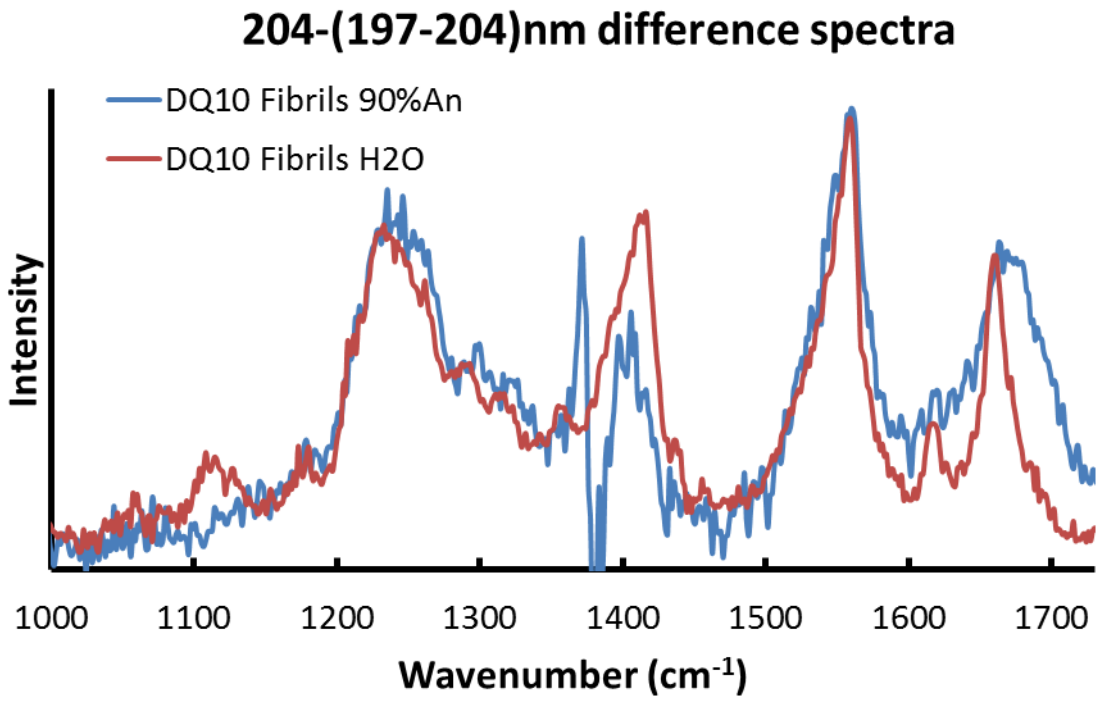

Figure A.9: Comparison of the 204-(197-204) nm difference spectra of DQ10 fibrils prepared in (red) water[46] and (blue) 90% acetonitrile. The 204-(197-204) nm difference spectrum highlights secondary amide vibrations of the peptide backbone. The AmI^S bands are at a different frequency for the two types of fibrils (~1660 cm^{-1} for fibrils prepared in water and at ~1670 cm^{-1} for fibrils grown in acetonitrile). Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

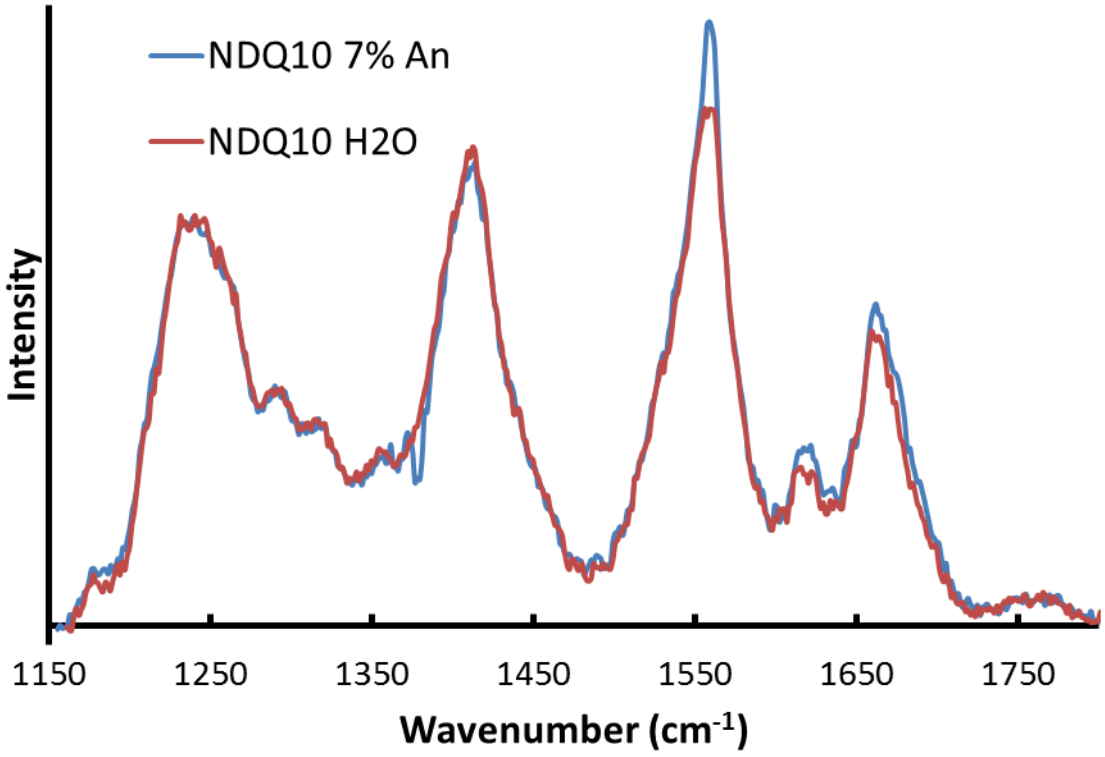

Figure A.10: Reversibility of acetonitrile induced β-sheet structure in NDQ10. UVRR spectra (204 nm) of NDQ10 in 100% H_2O (red) and NDQ10 in 93% H_2O 7% acetonitrile after dilution from 70% acetonitrile solution. The spectra appear essentially identical. This demonstrates the reversibility of the formation of low molecular weight β-sheet structures in NDQ10 at high acetonitrile concentrations. Reprinted with permission from Punihaole et al. *J. Phys. Chem. B.* 2017, 121 (24), 5953–5967. Copyright 2017 American Chemical Society.

Appendix B

Supporting Information for chapter 6.0

B.1 Estimation of Error in ΔH_{int} Measurements

To estimate the error of our experimental ΔH_{int} measurements, we estimate the error of the AmI^P frequency measurements and its propagation when converting AmI^P to ΔH_{int}. In Chapter 6.0, we use eq. B.1 to calculate ΔH_{int} from the AmI^P frequency of formamide:

$$AmI^P = \big(b + m(-40\ acc\#)\big) + (-18\ kcal^{-1}\ mol \cdot\ m\,\Delta H_{int}) \qquad \textbf{B.1}$$

where the parameters in eq. B.1 are the same as those described for eq. 6.5. This equation can be re-arranged to calculate ΔH_{int}:

$$\Delta H_{int} = \frac{AmI^P - (b + m(AN_{vac})}{(CF)m} \qquad \textbf{B.2}$$

where CF and AN_{vac} denotes the AN to ΔH_{int} conversion factor and AN of the "vacuum" solvent, respectively. The error in eq. propagates as:[220]

$$\sigma_{\Delta H} = |\Delta H_{int}|$$

$$\sqrt{\left(\frac{\sqrt{{\sigma_{AmI}}^2 + \left(\sqrt{\sigma_b^2 + |AN_{vac}(m)|\sqrt{\left(\frac{\sigma_m}{m}\right)^2 + \left(\frac{\sigma_{ANvac}}{AN_{vac}}\right)^2}^2}\right)^2}}{AmI - \left(b + m(AN_{vac})\right)}\right)^2 + \left(\frac{|CF(m)|\sqrt{\left(\frac{\sigma_{CF}}{CF}\right)^2 + \left(\frac{\sigma_m}{m}\right)^2}}{CF(m)}\right)^2} \qquad \textbf{B.3}$$

where σ denotes the error of the subscripted variable. For formamide, we determined the error of b (±2.82) and m (±0.092) for the linear fit to the formamide AmI^P-AN dependence (eq. 6.1). Cutmore and Hallam[154] report an error of ~±1.5 cm^{-1} for the determination of their band frequencies (σ_{AmI}).

Determining Error in the AN of Vacuum and the AN-ΔH_{int} Conversion Factor

As described in Chapter 6.0, AN_{vac} was found by equating eqs. 6.1 and 6.2:

$$AN_{vacc} = \frac{b' - b}{m} \qquad \textbf{B.4}$$

where *b'* is the slope of eq. 6.2. To determine the error in AN_{vac}, the error propagates as:[220]

$$\sigma_{ANvac} = |AN_{vac}|\sqrt{(\frac{\sqrt{{\sigma_{b'}}^2 + {\sigma_b}^2}}{b' - b})^2 + (\frac{\sigma_m}{m})^2} \qquad \textbf{B.5}$$

where the parameters in eq. B.5 are the same as that in the above equations, and σ is the error of the subscripted variable (σ_b = ±2.82, $\sigma_{b'}$ = ±5.09, and σ_m = ±0.092) for formamide. From eq. B.5, we find that σ_{ANvac} = ±13.76.

We also must determine the error of the conversion factor (eq. 6.3). The error propagates as:[220]

$$\sigma_{CF} \approx |-18\ acc\#\ \ kcal^{-1}\ mol\ | \sqrt{\left(\frac{\sigma_m}{-0.50\ cm^{-1} acc\#^{-1}}\right)^2 + \left(\frac{\sigma_{m'}}{8.8\ cm^{-1} kcal^{-1}\ mol}\right)^2} \qquad \textbf{B.6}$$

where σ_{CF} is the error in the conversion factor, σ_m (±0.0915 cm^{-1} $acc\#^{-1}$) is the error of the slope of eq.6.1, and $\sigma_{m'}$ (±1.39 cm^{-1} $kcal^{-1}$ mol) is the error of the slope of eq. 6.2. From this, we calculate an error of ±4.346 acc# $kcal^{-1}$ mol for CF.

Estimation of Error for Formamide ΔH_{int} Values

We have the uncertainty of each parameter in eq. and can calculate the error in ΔH_{int} from our AmI^P-AN dependence of formamide. The variables are as follows: m= -0.5±0.092 cm^{-1} $acc\#^{-1}$, b=1715±2.82 cm^{-1}, AmI= observed frequency±1.5 cm^{-1}, AN_{vac}= -40±13.76 acc#, CF=-18±4.346 acc# $kcal^{-1}$ mol, and ΔH_{int} is determined by eq. B.7 from a given AmI^P frequency.

$$\Delta H_{int} = \frac{AmI^P - (b + m(AN_{vac})}{(CF)m} = \frac{AmI^p - 1735\ cm^{-1}}{9\ cm^{-1} kcal^{-1} mol} \qquad \textbf{B.7}$$

Thus, ΔH_{int} and its error depend on the experimental AmI^P frequency. From eq. we calculate the error in ΔH_{int} for formamide for the average AmI frequency observed (1702 cm^{-1}) and find an error of ~±1.45 kcal mol^{-1}.

Estimation of Error for Q10 ΔH_{int} Values

For our calculation of ΔH_{int} of Q10 however, we use the AmI-AN dependence of propanamide instead of formamide because propanamide better models the Q side chain. Because we do not know the true AmI^P-AN or AmI^P-ΔH_{int} dependence of Q10, we cannot obtain the errors in the slopes and intercepts needed to directly determine our error in calculating Q10 ΔH_{int}.

Instead, to estimate our errors for ΔH_{int} of Q10, we assume that the correlations for Q10 will be similar to that of formamide. However, we now use the slope (m_p=-0.64 cm^{-1} $acc\#^{-1}$) and intercept (b_p=1704 cm^{-1}) of the dependence of propanamide AmI^P band on AN. Also, in Chapter 6.0, we assume that the AN-ΔH_{int} conversion factor (CF) and the AN of vacuum (AN_{vac}) are constant, and we will also assume this to be true for our error estimation. Because the values of m and b are different for propanamide, we use eqs. B.5 and B.6 to recalculate the error in our AN_{vac} and CF values. We find the errors to be ±10.6 acc# and ±3.38 acc# $kcal^{-1}$ mol for AN_{vac} and CF respectively. In our calculation of AN_{vac} and CF errors, we use the equations b'=b+m(AN_{vac}) and m'=m(CF) from Chapter 6.0 to determine the value of b' and m' respectively.

We now have every variable needed to estimate the error in calculating ΔH_{int} for Q10 from the AmI^P-AN dependence of propanamide: m_p= -0.64±0.092 cm^{-1} $acc\#^{-1}$, b_p=1704±5.09

cm^{-1}, AmI= observed frequency±1.5 cm^{-1}, AN_{vac}= -40±10.64 acc#, CF=-18±3.38 acc# $kcal^{-1}$ mol, and ΔH_{int} is determined by eq. B.8 for a given AmI^P frequency.

$$\Delta H_{int} = \frac{AmI^P - (b + m(AN_{vac})}{(CF)m} = \frac{AmI^p - 1730\ cm^{-1}}{11.52\ cm^{-1}kcal^{-1}mol} \qquad \textbf{B.8}$$

ΔH_{int} and its error depend on the experimental AmI^P frequency which ranges from ~1660 to ~1680 cm^{-1} for Q10. From eq. , we estimate an error of ~±1.25 kcal mol^{-1} and ~±1.59 kcal mol^{-1} for Q10 side chains hydrated (1680 cm^{-1}) and in peptide-peptide interactions (1660 cm^{-1}), respectively. Our estimated error is less than the ~1.7 kcal mol^{-1} difference in ΔH_{int} between solvated side chains and side chains involved in peptide-peptide interactions. In contrast, our estimated error is much greater than the ~0.5 kcal mol^{-1} difference in ΔH_{int} between DQ10 and NDQ10 fibrils.

Error in Abscissa of Equation 6.1 (AN) Negligible

It is important to note that our linear regression for the dependence of the AmI^P frequency on AN (eq. 6.1) and AmI^P frequency on ΔH_{int} (eq. 6.2) assumes that there is negligible error in the solvent AN and ΔH_{int} values, respectively. Here, we show that this assumption is valid by comparing the errors of the measured AN and ΔH_{int} values to that of the AmI^P frequency.

Solvent AN values were determined via NMR measurements of triethylphosphine (Et_3OP) by Mayer et al.[156] The ^{31}P chemical shift of Et_3PO was measured with an error of <±0.05 ppm. The AN value is calculated from the chemical shift using the following formula:[156]

$$AN = \frac{\delta_{corr}}{\delta_{corr}(SbCl_5 \cdot Et_3PO)} \times 100 \quad \textbf{B.9}$$

where $\delta_{corr}(SbCl_5 \cdot Et_3PO)$ is the ^{31}P chemical shift of Et_3PO in $SbCl_5$ at infinite dilution and δ_{corr} is the chemical shift of Et_3PO in the solvent of interest. Thus, the error in chemical shift propagates as:[220]

$$\sigma_{AN} \approx \left|\frac{\delta_{corr}}{\delta_{SbCl5}}\right| \sqrt{\left(\frac{\sigma_{corr}}{\delta_{corr}}\right)^2 + \left(\frac{\sigma_{SbCl_5}}{\delta_{SbCl_5}}\right)^2} \times 100 \quad \textbf{B.10}$$

where σ_{AN} is the standard deviation of the AN value, δ_{corr} is the ^{31}P chemical shift of Et_3PO in a solvent of interest, δ_{SbCl5} is the ^{31}P chemical shift of Et_3PO in $SbCl_5$, σ_{corr} is the standard deviation of δ_{corr}, and σ_{SbCl5} is the standard deviation of δ_{SbCl5}. Using an error of ±0.05 ppm, chemical shift values, and AN values reported by Mayer et al.[156] we calculate that σ_{AN} is ~±0.12.

In comparison, Cutmore and Hallam[154] report a "precision" of ~±1.5 cm^{-1} for their band frequency measurements. To directly compare σ_{AN} and σ_{AmI}, we convert σ_{AN} to σ_{AmI} using eq.

6.1, neglecting error in the slope of the linear fit. The error propagates as:

$$\sigma_{AmI} = \sigma_{AN}|-0.5\ cm^{-1}\ acc\#^{-1}| \quad \textbf{B.11}$$

An error of σ_{AN}~±0.12 corresponds to an error of σ_{AmI}~±0.06. Therefore, the error introduced to the correlation by the measurement of AN (σ_{AmI}~±0.06) is ~4% of that introduced by the measurement of the AmI frequency (σ_{AmI}~±1.5 cm^{-1}), and we can neglect the error in the measurement of the solvent AN.

Error in Abscissa of Equation 6.2 (ΔH_{int}) Negligible

Thermodynamic data on the solvation of formamide in different solvents was collected by Varfolomeev et al.[153] The enthalpy of solvation (ΔH_{solv}) was determined for the dissolution of formamide in different solvents with an error of ~±0.05 kJ mol^{-1} (~±0.012 kcal mol^{-1}) and this value was used to calculate the enthalpy of hydrogen bonding (ΔH_{int}) between formamide and various solvents. We were unable to explicitly determine the error for the ΔH_{int} values reported by Varfolomeev et al.[153]. Therefore, we use the ΔH_{solv} error reported by Varfolomeev et al.[153] as an estimate of the error of the ΔH_{int} values they report for formamide.

To directly compare $\sigma_{\Delta H}$ and σ_{AmI}, we convert $\sigma_{\Delta H}$ to σ_{AmI} using eq. 6.2. Neglecting error in the slope of the linear fit, the error propagates as:[220]

$$\sigma_{AmI} = \sigma_{\Delta H} |8.8\ cm^{-1}\ kcal^{-1} mol| \qquad \textbf{B.12}$$

An error of $\sigma_{\Delta H}$~±0.012 kcal mol^{-1} corresponds to an error of σ_{AmI}~±0.1 cm^{-1}. Therefore, the error introduced to the correlation by the measurement of ΔH (σ_{AmI}~±0.1 cm^{-1}) is ~7% of that introduced by the measurement of the AmI frequency (σ_{AmI}~±1.5 cm^{-1}), and we can neglect the error in the measurement of the solvent ΔH.

B.2 **Fitting the AmI^{s} Spectra of Q10 Fibrils**

Figure B.1 shows our fitting of the AmI^{S} UVRR spectra for Q10 fibrils prepared from β-strand and PPII monomers. The data was originally reported by Punihaole et al.[46] The spectra were fit using a script written in MATLAB (ver. 2015b, Natick, MA, MathWorks Inc.) . The spectra, I(ν), were best-fit to the sum of four Lorentzian bands:

$$I(\upsilon) = \sum_{i=1}^{n=4} \frac{H_i}{4\left(\frac{\upsilon - \upsilon_i}{w_i}\right)^2 + 1} \qquad \textbf{B.13}$$

where H_i is the amplitude, ν_i is the center frequency, and w_i is the width of the i^{th} Lorenztian band.

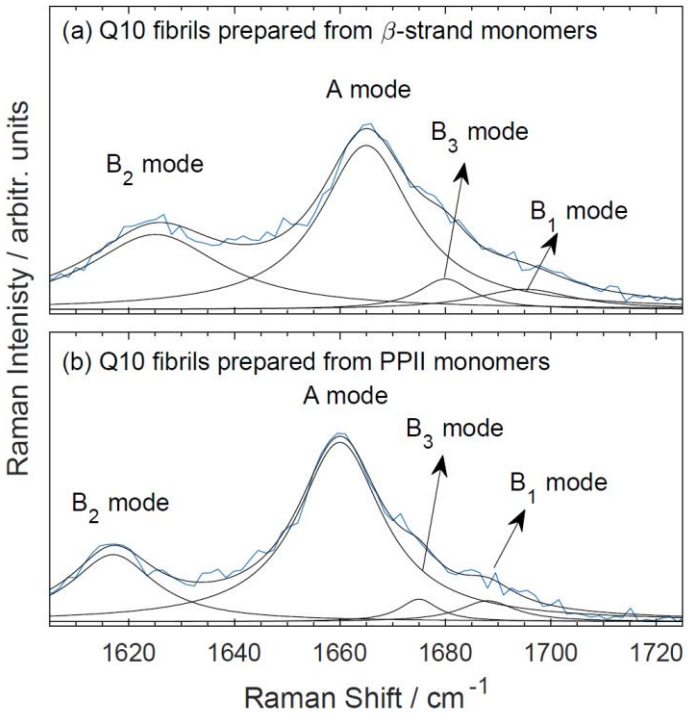

Figure B.1: Fitting of the AmI^S region of the UVRR spectra of Q10 fibrils prepared from (a) β-strand and (b) PPII monomers. The spectra were taken from data collected by Punihaole et al.46 The spectra were fit to the sum of four Lorentzian bands. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

Appendix C

Supporting Information for Chapter 7.0

C.1 UVRR Curve-Fitting

Grams software (version 8.0, Thermo Fisher Scientific, Inc.) was used to model the Q20 UVRR spectra as a sum of Gaussian and Lorentzian bands:

$$\boldsymbol{f(x) = \sum_i (1 - M)(He^{-(\frac{x-x_0}{w})^2(4\ln(2))}) + (M)\left(\frac{H}{4(\frac{x-x_0}{w})^2+1}\right)} \qquad \textbf{C.1}$$

where H is the peak height, x_0 is the center peak frequency, w is the full width at half height, M=1 if the band is Lorentzian, and M=0 if the band is Gaussian for the i^{th} UVRR band. The $AmIII_3^S$ bands where modeled as Gaussians because they are inhomogenously broadened by the Ψ angle distribution of the peptide.[74]

C.2 Ramachandran Ψ Angle Calculation

Methodologies for determining the relationship between the $AmIII_3^S$ frequency and Ψ angle are reported by Mikhonin et al.[2] The procedure used to determine the Ψ angle distribution from the inhomogeneous width of the $AmIII_3^S$ band was previously reported by Asher et al.[1] We

assume that the inhomogeneous broadening of the $AmIII_3^S$ band results from only the Ψ angle distribution of the peptide. The Gaussian $AmIII_3^S$ bands where modeled as a sum of Lorentzian bands with identical homogenous linewidths:

$$AmIII_3^S = \frac{1}{\pi}\sum_{i}^{M}\frac{p_i\Gamma^2}{\Gamma^2 + (v - v_i)^2} \qquad \textbf{C.2}$$

where p_i is the probability that the i^{th} band occurs at center frequency ν_i and Γ is the homogenous linewidth of the $AmIII_3^S$ band. The $AmIII_3^S$ homogenous linewidth was previously estimated to be ~7.5 cm^{-1} for small crystalline peptides.[74] The Ψ angle for each Lorentzian band was calculated using the appropriate equation derived by Mikhonin et al.[64] to obtain a Ψ angle distribution.

The following equation was developed by Mikhonin et al.[64] to calculate Ψ from the $AmIII_3^S$ frequency for a peptide backbone completely solvated by water:

$$\boldsymbol{v_i = 1256\ (cm^{-1}) - 54(cm^{-1})\sin(\Psi + 26) - 0.11(\frac{cm^{-1}}{^{\circ}C})(T - T_0)} \qquad \textbf{C.3}$$

where ν_i is the frequency of the i^{th} Lorentzian band, T is the sample temperature (°C), and T_0 = 0^0C. We use this equation to calculate Ψ angles for the PPII and 2.5_1-helix conformations of DQ20 and the NDQ20 supernatant, which have peptide backbones that are completely solvated in water.

The following equation is used to calculate the Ψ angle distribution for a peptide backbone in an unknown solvation environment:

$$\boldsymbol{v_i = 1250\ (cm^{-1}) - 54(cm^{-1})\sin(\Psi + 26) - 0.06\left(\frac{cm^{-1}}{^{\circ}C}\right)(T - T_0)}$$

C.4

where the variables are the same as discussed above for equation C.3. This equation was used to calculate the Ψ angle distribution for the turn-like structures found in DQ20 where the peptide backbone can engage in either peptide-peptide and/or peptide-water hydrogen bonds.

To calculate the Ψ angle of Q20 powder and NDQ20 in water, we used the following equation by Mikhonin et al.[64] to calculate Ψ for crystalline peptides:

$$\boldsymbol{v_i = 1260\ (cm^{-1}) - 54(cm^{-1})\sin(\Psi + 26)} \qquad \textbf{C.5}$$

This equation is used to calculate the Ψ angle for solid, anhydrous peptides. To calculate Ψ for NDQ20 fibrils, we use the equation developed by Mikhonin et al.[64] for anhydrous β-sheet and α-helix conformations.

$$\boldsymbol{v_i = 1239\ (cm^{-1}) - 54(cm^{-1})\sin(\Psi + 26)} \qquad \textbf{C.6}$$

This equation is used for anhydrous peptide backbones involved peptide-peptide hydrogen bonding of canonical β-sheet and/or α-helical structures.

C.3 **Absorbance of TFA**

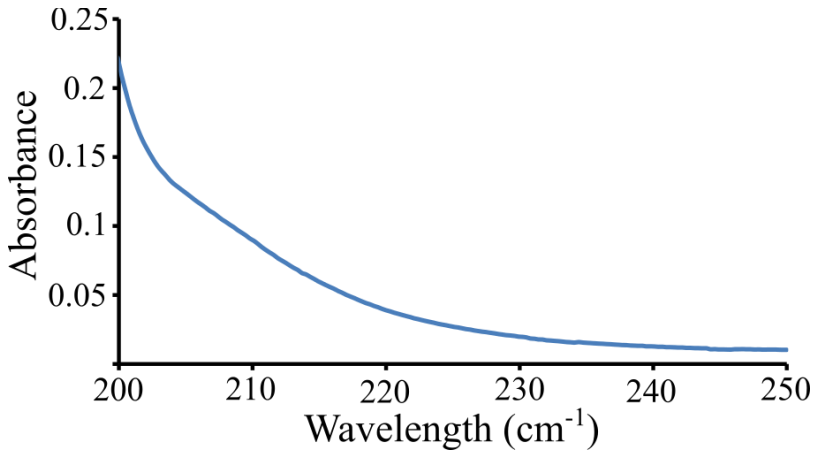

Figure C.1: Absorbance spectrum of 1% (v/v) TFA in water. Reprinted with permission from Jakubek et al. ***J. Phys. Chem. B.*****, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.**

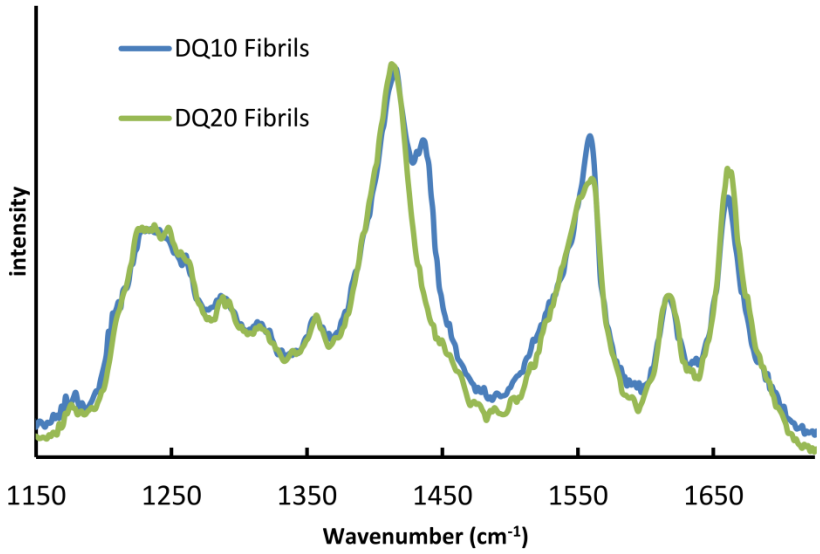

Figure C.2: Comparison of 204 nm UVRR spectra of (blue) DQ10 fibrils and (green) DQ20 fibrils. The $AmIII_3^S$ and AmI spectral regions are similar indicating that DQ10 and DQ20 fibrils have similar secondary structure and hydrogen bonding interactions. The only significant difference between the two spectra is the presence of a strong band at ~1435 cm^{-1} in DQ10 fibrils that is not observed in DQ20 fibrils. This band has the same frequency as the C=O stretching band of TFA, which is strongly resonance enhanced at 204 nm excitation (see Chapter 7.0). It is possible that this band arises from TFA in the supernatant of the DQ10 fibrils. When the DQ20 fibrils were prepared they were washed three times with nanopure H_2O to remove any TFA in the supernatant. Reprinted with permission from Jakubek et al. *J. Phys. Chem. B.*, 2019, 123, 8, 1749-1763. Copyright 2019 American Chemical Society.

Appendix D

Supporting Information for Chapter 8.0

D.1 UVRR Curve-Fitting

We used Grams software (version 8.0, Thermo Fisher Scientific, Inc. Waltham, Mass., USA) to model each UVRR spectrum as a sum of Gaussian and Lorentzian bands:

$$f(x) = \sum_i (1 - M)(He^{-(\frac{x-x_0}{w})^2(4\ln(2))}) + (M)\left(\frac{H}{4(\frac{x-x_0}{w})^2+1}\right) \quad \textbf{D.1}$$

where H is the peak height, X_0 is the peak frequency, w is the full width at half height, M=1 if the band is Lorentzian, and M=0 if the band is Gaussian for the i^{th} UVRR band. The $AmIII_3^S$ fitted bands were modeled as Gaussian bands because they are inhomogenously broadened due to the peptide Ψ angle distribution.[74]

D.2 Ψ Ramachandran Angle Calculation

Details describing the methodologies for calculating Ψ angle distributions are reported by Mikhonin et al.[64,74] We assume that the inhomogeneous broadening of the $AmIII_3^S$ band results

from only the Ψ angle distribution of the peptide. The Gaussian $AmIII_3^S$ bands were modeled as a sum of Lorentzian bands with identical homogenous linewidths:

$$AmIII_3^S = \frac{1}{\pi}\sum_i^M \frac{p_i \Gamma^2}{\Gamma^2+(v-v_i)^2} \qquad \textbf{D.2}$$

where p_i is the probability that the i^{th} band occurs at center frequency υ_i, and Γ is the homogenous linewidth of the $AmIII_3^S$ band. The $AmIII_3^S$ homogenous linewidth was previously estimated from peptide crystals to be ~7.5 cm^{-1}.[74] After fitting the $AmIII_3^S$ band to Lorentzian bands, we calculated the Ψ angle for each Lorentzian band using the equations derived by Mikhonin et al.[64] to obtain a Ψ angle distribution.

We use the following equation to calculate the Ψ angle distribution for the PPII-like and 2.5_1-helix-like $AmIII_S^3$ bands of DQ peptides:

$$v_i = 1256\ (cm^{-1}) - 54(cm^{-1})\sin(\Psi + 26) - 0.11(\frac{cm^{-1}}{^\circ C})(T - T_0) \qquad \textbf{D.3}$$

where υ_i is the frequency of the i^{th} Lorentzian band, T=18°C is the sample temperature, and T_0=0°C. As discussed by Mikhonin et al.,[64] this equation is used when the peptide bonds are fully exposed for water as expected for PPII and 2.5_1-helix conformations.

We use the following equation to calculate the Ψ angle distribution for the turn-like and β-strand-like $AmIII_3^S$ bands of DQ and NDQ peptides respectively:

$$v_i = 1250\ (cm^{-1}) - 54(cm^{-1})\sin(\Psi + 26) - 0.06(\frac{cm^{-1}}{^\circ C})(T - T_0) \qquad \textbf{D.4}$$

where the variables are the same as those defined for equation D.3. As discussed by Mikhonin et al.,[64] this equation is used when the hydration state of the peptide backbone is unknown.

D.3 **Modeling DQ15 and DQ20 CD Spectra**

We modeled the CD spectra of DQ15 and DQ20 as a linear sum of DQ10 and NDQ15 basis spectra. In the Chapter 8.0 we show that NDQ10-15 peptides are in a predominately β-strand-like conformation with no noticeable differences in structure. Therefore, either the NDQ10 or NDQ15 CD spectrum could be used as a CD basis spectrum for the pure β-strand-like polyQ conformation. We use the CD spectrum of NDQ15 as our β-strand-like basis spectrum. In the Chapter 8.0 we also show that longer DQ peptides contain a larger sub-population of β-strand conformation. Therefore, we use our smallest DQ peptide (DQ10) as our basis spectrum for the pure DQ PPII-like structure. Spectra were modeled using the following equation:

$$Model\ Spectrum = \left(DQ10\ spectrum\ \times C_{DQ10}\right) + \left(NDQ15\ spectrum\ \times C_{NDQ15}\right) \quad \textbf{D.5}$$

where C_{DQ10} and C_{NDQ15} are coefficients. The coefficients were adjusted to minimize the sum of the residual between the spectrum of interest and the model spectrum. We estimated the relative amount of PPII-like and β-strand-like structure in our model spectra from the coefficient values as follows:

$$\%\beta strand = \frac{C_{NDQ15}}{C_{NDQ15}+C_{DQ10}} \quad \textbf{D.6}$$

$$\%PPII = \frac{C_{DQ10}}{C_{NDQ15}+C_{DQ10}} \quad \textbf{D.7}$$

Our estimates assume that the DQ10 and NDQ15 CD basis spectra represent the pure PPII-like and β-strand-like structures of the polyQ peptide, respectively.

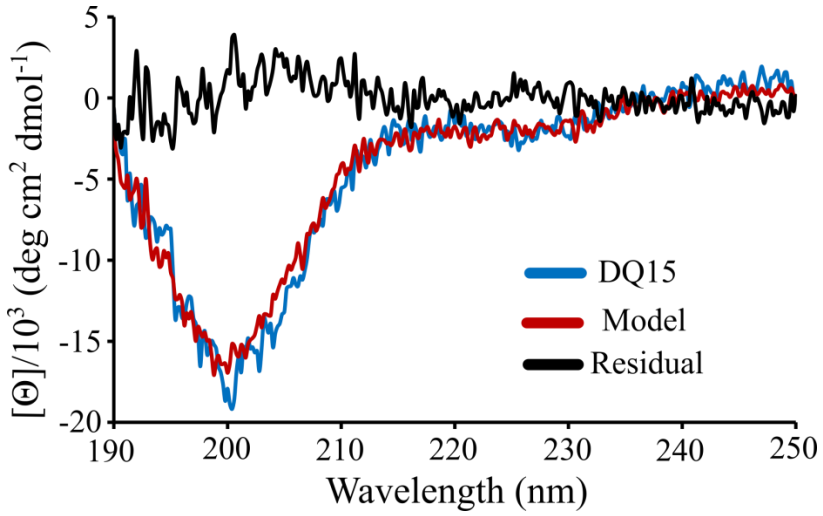

Figure D.1: CD spectral modeling of DQ15. (Blue) CD spectrum of DQ15, (Red) model spectrum from linear sum of DQ10 and NDQ15 spectra, (Black) residual (DQ15-Model difference spectrum. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019 123(19), 4193-4203. Copyright 2019 American Chemical Society.

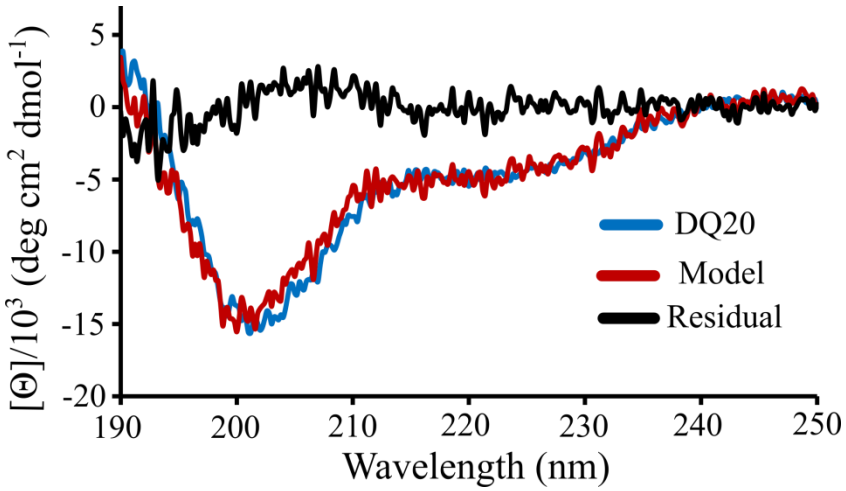

Figure D.2: CD spectral modeling of DQ20. (Blue) CD spectrum of DQ20, (Red) model spectrum from linear sum of DQ10 and NDQ15 spectra, (Black) residual (DQ20-Model difference spectrum). Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

D.4 **Modeling β-Strand Sub-populations in DQ Peptides**

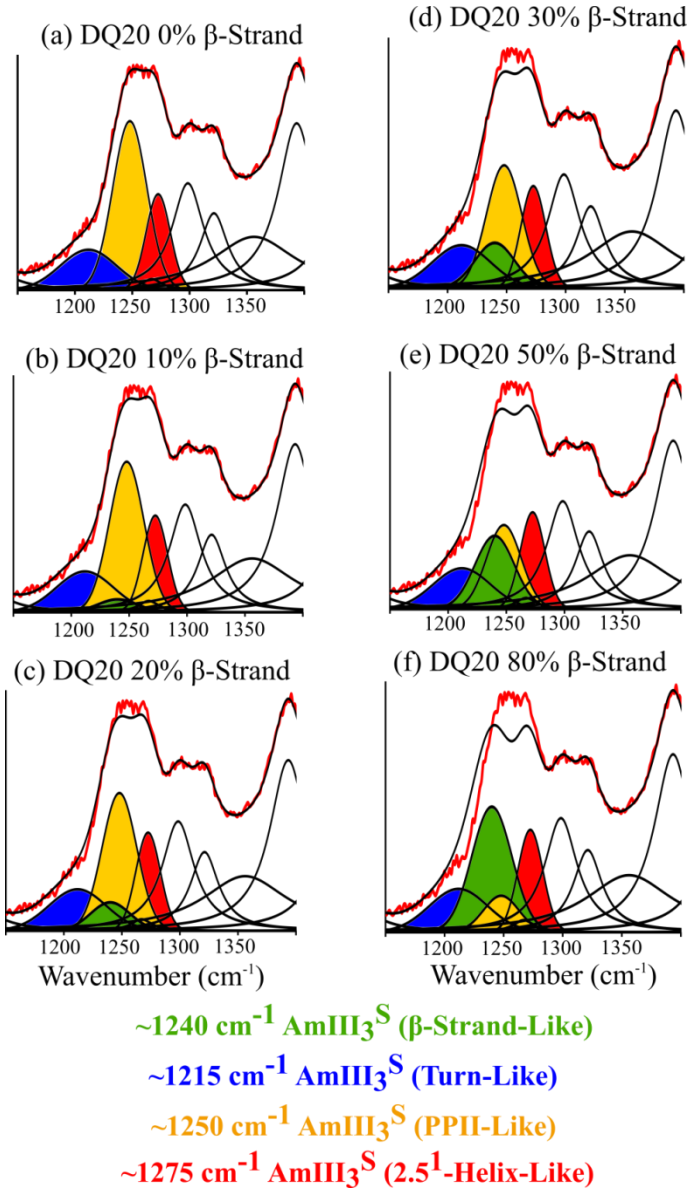

Figure D.3: Spectra of DQ20 with 1240 cm^{-1} band artificially added to model (a) 0%, (b) 10%, (c) 20%, (d) 30%, (e) 50%, and (f) 80% sub-populations of β-strand-like structure in DQ peptides. The 1250 cm^{-1} band (yellow) is that of the PPII-like conformation found in DQ peptides. The 1240 cm^{-1} band (green) is artificially added to model a β-strand sub-population. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

D.5 **Temperature Dependence of DQ15**

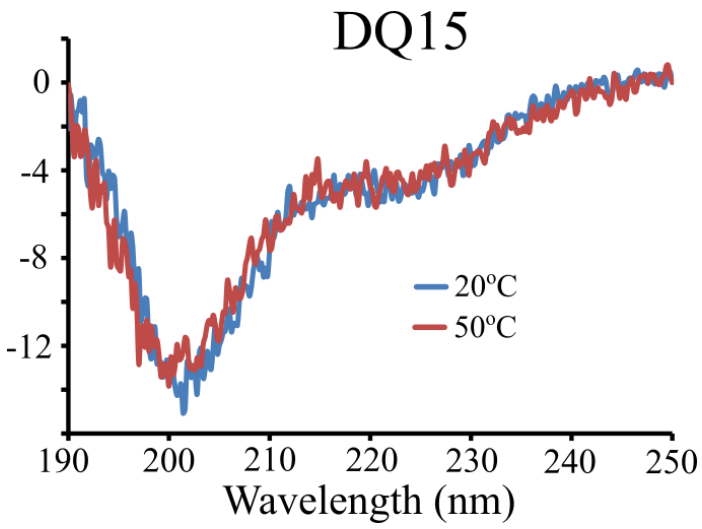

Figure D.4: Temperature dependence of the DQ15 CD spectrum. The DQ15 CD spectra at 20°C and 50°C are similar indicating only minute structural changes. In contrast, DQ20 shows a significant structural change upon heating from 20°C to 50°C (Figure 8.10). This difference occurs because DQ15 contains a smaller β-strand-like sub-population compared to DQ20. As discussed in Chapter 8.0, the β-strand-like structure is sensitive to a temperature increase from 20°C to 50°C while the PPII-like structure shows negligible change. Reprinted with permission from Jakubek et al., *J. Phys. Chem. B.*, 2019, 123(19), 4193-4203. Copyright 2019 American Chemical Society.

Appendix E

Supporting Information for Chapter 9.0

E.1 DOSY NMR

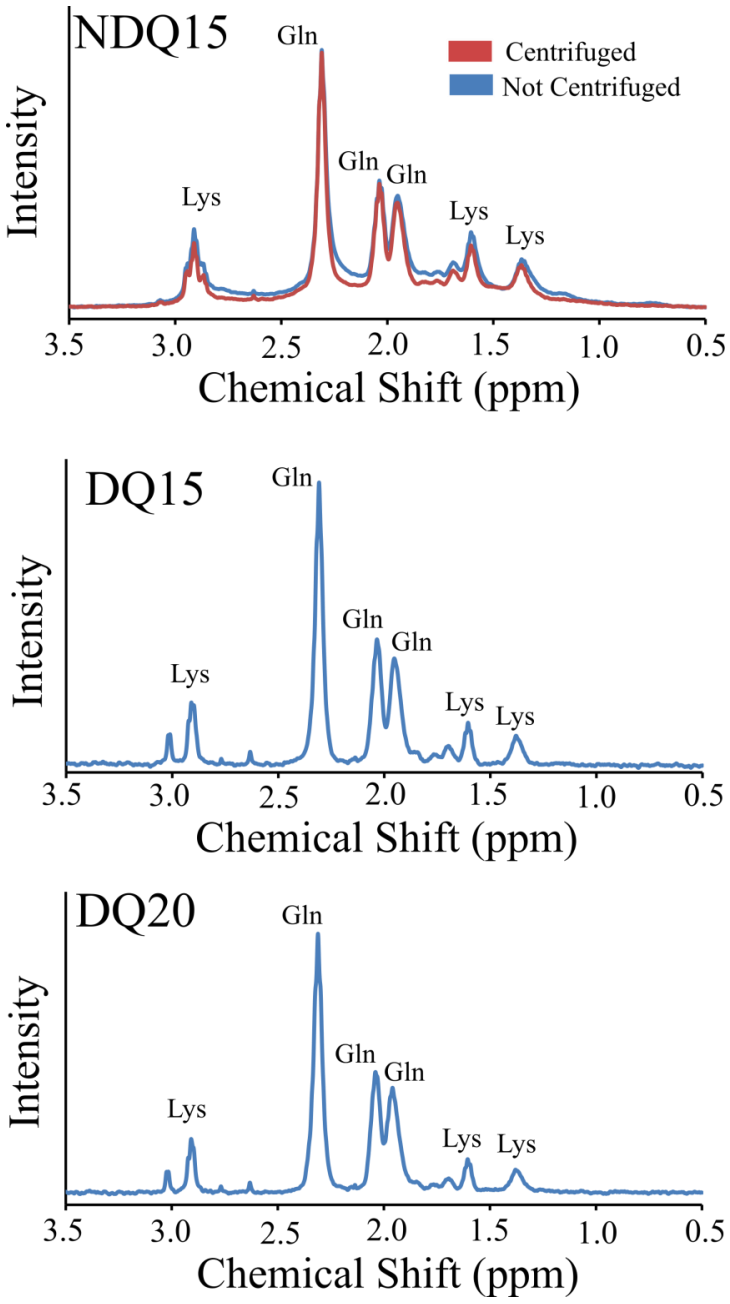

Figure E.1: ^{1}H NMR spectra of NDQ15, DQ15, and DQ20. Peaks originating from protons on Gln and Lys residues are labeled.

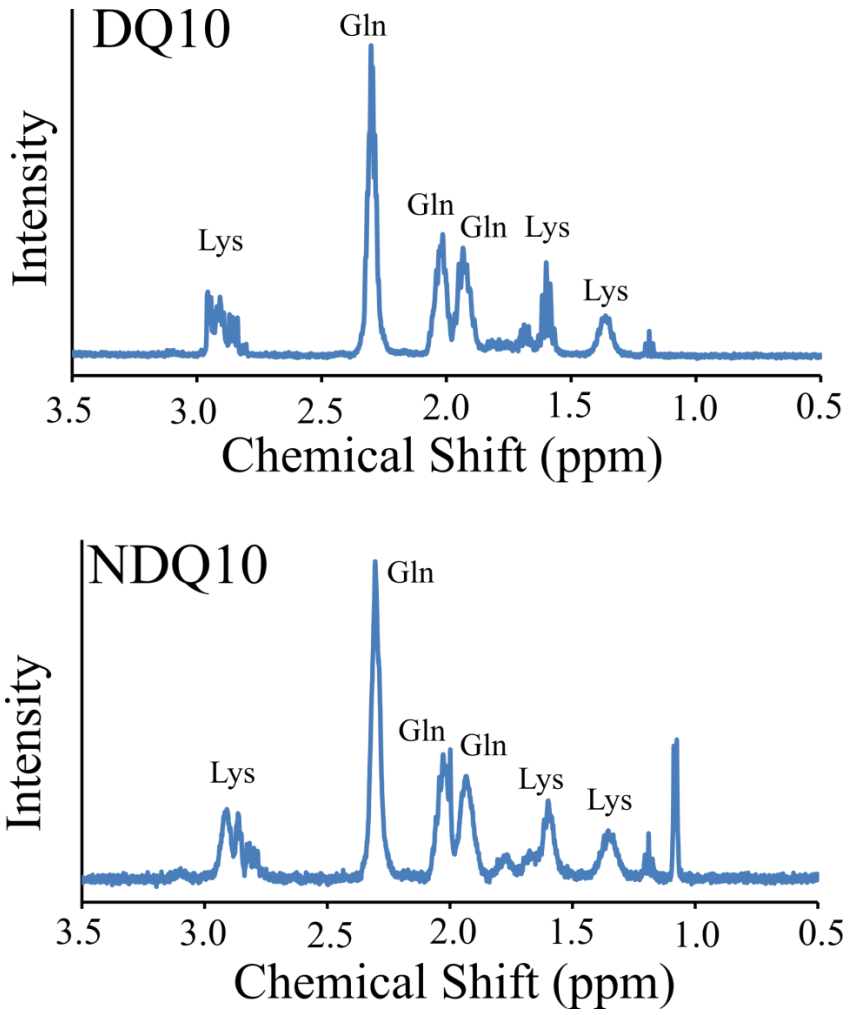

Figure E.2: ^{1}H NMR spectra of DQ10 and NDQ10 previously collected by Punihaole et al.[71] Peaks originating from protons on Gln and Lys residues are labeled.

^{1}H NMR chemical shifts where assigned using the work of Bundi et al.[221] The peaks found at ~2.00 ppm are assigned to the $C_{\beta}H$ hydrogens of Gln, and the peak at ~2.35 ppm is assigned to the $C_{\gamma}H$ hydrogen of the Gln side chain. The peaks found at ~1.4 ppm, ~1.6 ppm, and ~2.9 ppm are assigned to $C_{\gamma}H$, $C_{\delta}H$, and $C_{\varepsilon}H$ hydrogens of the Lys side chain, respectively.

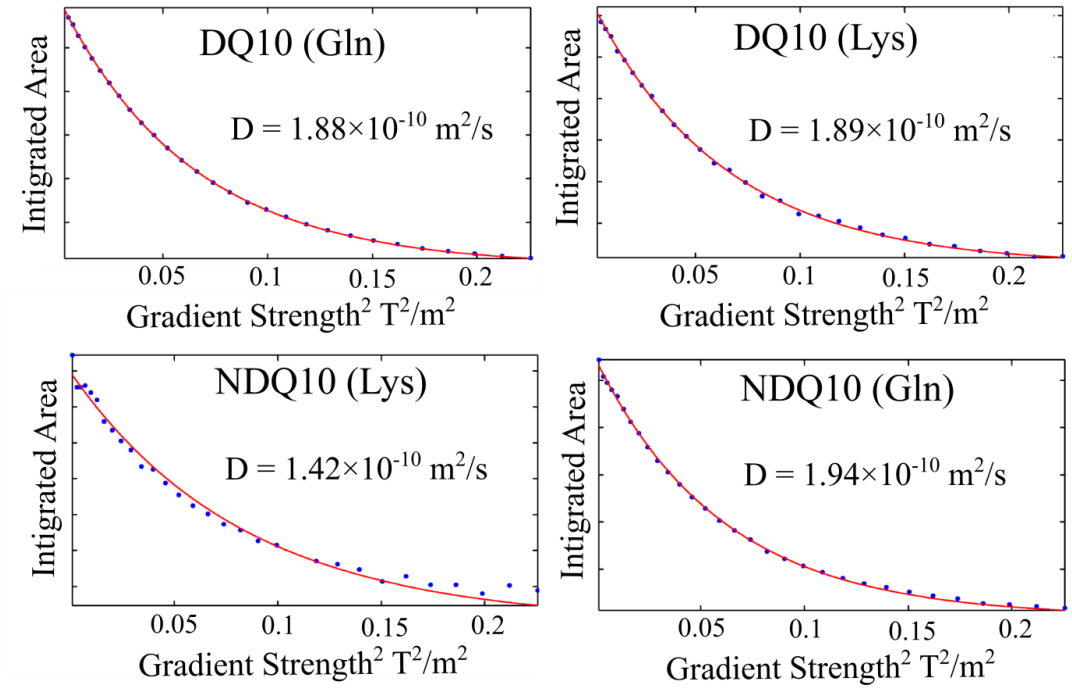

Figure E.3: DOSY NMR signal decays for the Glutamine and Lysine resonances of DQ10 and NDQ10.

E.2 **Diffusion Coefficient and Van Der Waals Volume of DQ Peptides**

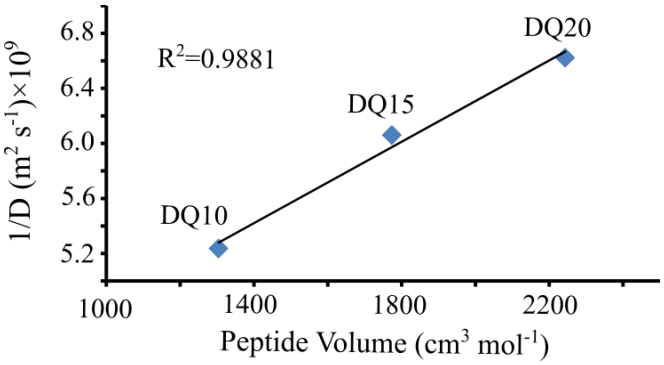

Figure E.4: Linear correlation between the reciprocal of the diffusion coefficient and molecular volume for DQ peptides. Molecular volumes of DQ10, DQ15, and DQ20 were calculated by summing the volumes of Gln, Lys, and Asp amino acids reported by Zamyatnin.[222]

E.3 **Time Dependence of NDQ15 NMR Spectra**

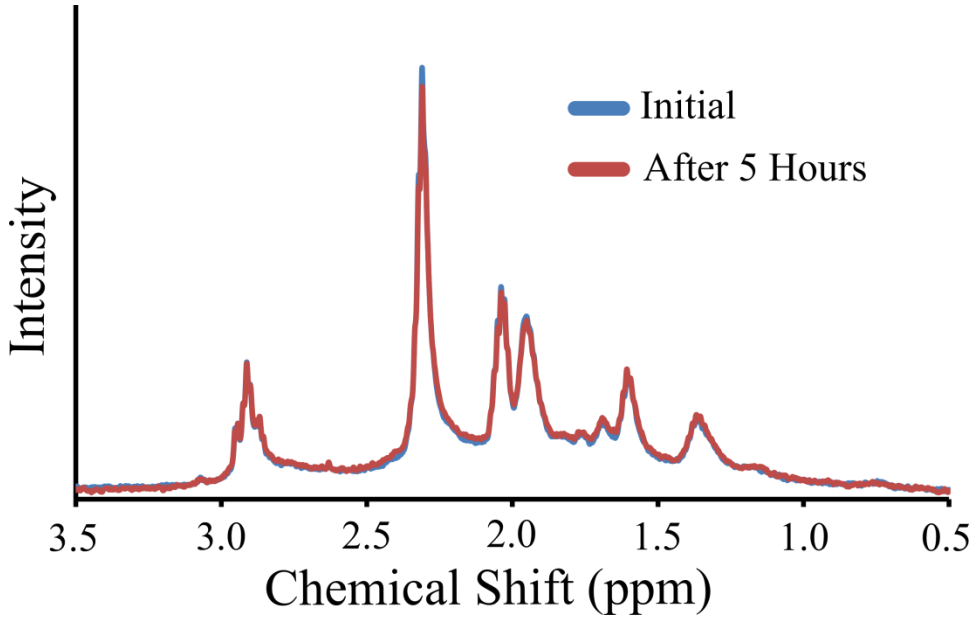

Figure E.5: H^1 NMR spectra of NDQ15 after sample preparation and 5 hours after sample preparation. Spectra are normalized to the Lys 1.3 ppm peak. The spectra are essentially identical indicating that the relative concentration of the monomer and oligomer species does not significantly change during DOSY data collection.

Appendix F

Description of Ψ Angle Calculation Procedure

F.1 Determination of Ψ Angle Distributions of the $AmIII_3^S$ band

The $AmIII_3^S$ (where the superscript "S" denotes the secondary amide peptide bond) Ψ dihedral angle distributions were calculated using methodologies previously described in detail.[62] First, the UVRR spectrum of a polyQ peptide between 1150 cm^{-1} and 1750 cm^{-1} was assigned and fit to a sum of Gaussian and Lorentzian bands based on our previous, detailed UVRR assignments of small peptides[62] and of Gln side chains.[71] The UVRR spectral fit and assignments of disaggregated Q20 (DQ20) and non-disaggregated Q15 (NDQ15) are shown in figure 1.

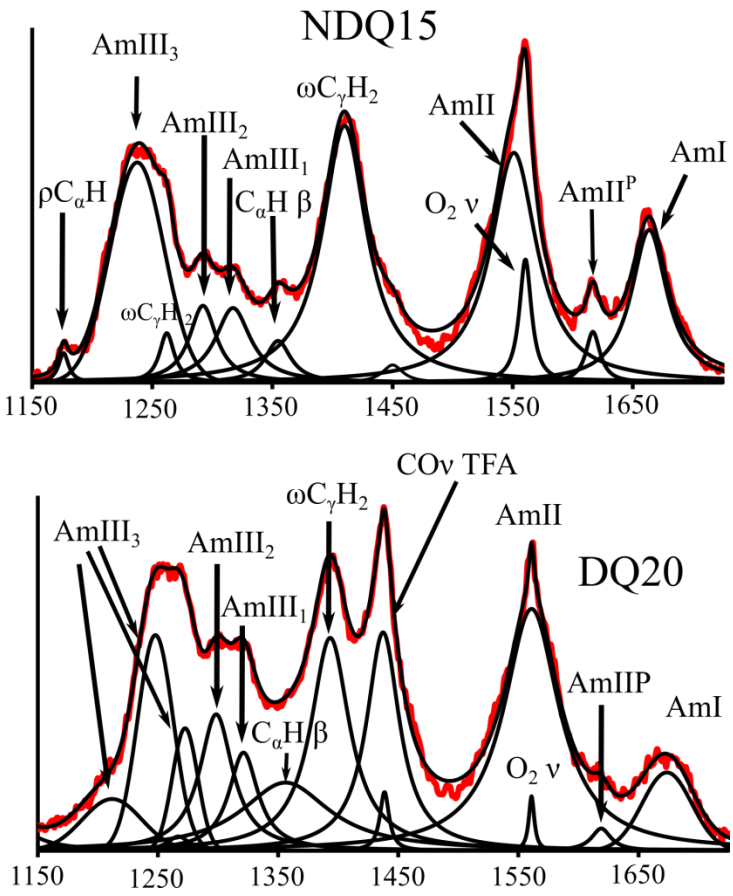

Figure F.1: UVRR (204 nm) spectral fits and assignments of NDQ15 and DQ20.

The bands found in the ~1650-1700 cm^{-1} spectral region are assigned to the AmI bands of the peptide backbone and Gln side chains.[62,118] The band at ~1615 cm^{-1} is assigned to the AmII band of the Gln side chains ($AmII^P$) based on our previous assignments of Gln.[71] The band at ~1550 cm^{-1} is assigned to the AmII band of the peptide backbone, which overlaps the atmospheric oxygen O_2 stretching band (~1560 cm^{-1}). The intense band at ~1400 cm^{-1} is assigned to the $C_\gamma H_2$ wagging band of the Gln side chain.[71] The intense band observed at ~1435 cm^{-1} in the spectrum of the disaggregated (but <u>not</u> of the non-disaggregated) peptides is assigned to the CO stretching band of trifluoroacetic acid (TFA) that is left over after the disaggregation procedure.[193] The bands at ~1355, ~1320, and ~1295 cm^{-1} are assigned to the $C_\alpha H$ bending, $AmIII_1$, and $AmIII_2$ bands, respectively.[82,118] The low intensity band at ~1265 cm^{-1} is assigned to the $C_\gamma H_2$ wagging band of the Gln side chain.[71] The band observed at ~1160 cm^{-1} is assigned to the $C_\alpha H$ rocking band of the Gln side chain.[71]

The $AmIII_3^S$ bands of peptides are found in the ~1200-1300 cm^{-1} spectral region. The $AmIII_3^S$ bandshape and frequencies depend on the peptide's secondary structure. Thus, the $AmIII_3^S$ bands of the disaggregated (DQ) and non-disaggregated (NDQ) polyQ peptides differ. For NDQ peptides we find a single $AmIII_3^S$ band located at ~1240 cm^{-1}. In contrast, DQ peptides have $AmIII_3^S$ bands located at ~1275, ~1250, and ~1215 cm^{-1}. The $AmIII_3^S$ bands were fit to Gaussians because the distribution of peptide Ψ angles must inhomogenously broaden the $AmIII_3^S$ bands.

We assume that the inhomogenously broadened $AmIII_3^S$ band ($B(\nu)$) can be modeled as the sum of M Lorentzian bands of identical homogeneous linewidth:

$$B(\nu) = \frac{1}{\pi} \sum_{i}^{M} \frac{p_i \Gamma^2}{\Gamma^2 + (\nu - \nu_i)^2} \qquad \textbf{F.1}$$

where p_i is the probability for the i^{th} Lorentzian band to occur at center frequency ν_i, while Γ is the homogeneous linewidth of the $AmIII_3^S$ vibration. We previously determined from peptide crystals that Γ is ~7.5 cm^{-1} for the $AmIII_3^S$ bands.[71,74] The $AmIII_3^S$ bands were fit to the sum of the fewest number of Lorentzian bands (M) needed to accurately model the $AmIII_3^S$ band shape. We utilized Γ =7.5 cm^{-1} and a 5 cm^{-1} separation between Lorentzians.

After fitting the inhomogeneously broadened $AmIII_3^S$ bands to a sum of Lorentzians, we calculated the individual contributing Ψ dihedral angles from the center frequency of each Lorentzian band as described in section F.2.

F.2 **Correlating the $AmIII_3^S$ Frequencies to their Ψ Angle Distribution**

We calculated the Ψ angle distributions from the set of $AmIII_3^S$ Lorentzian band frequencies using the equations of Mikhonin et al. who proposed different $AmIII_3^S$ ν_i-Ψ angle correlations for peptide bonds with different peptide bond solvation states.[64]

We calculated Ψ angle distributions for the $AmIII_3^S$ bands indicative of PPII and 2.5_1-helix structures using eq. F.2:

$$\upsilon_i = 1256\ \mathrm{cm^{-1}} - 54\mathrm{cm^{-1}}\sin(\Psi + 26) - 0.11\ \mathrm{cm^{-1}}/°C(T - T_0) \qquad \textbf{F.2}$$

where T = 18°C is the sample temperature and T_0 = 0°C. Rearrangement of this equation to solve for Ψ yields the equation:

$$\Psi = \arcsin\left(\frac{\upsilon_{AmIII} - 1256 + 0.11(T - T_0)}{-54}\right) - 26 \qquad \textbf{F.3}$$

As discussed in detail by Mikhonin et al.,[64] eq. F.3 is used for situations when the peptide bonds are fully exposed to water, such as in the case of PPII-like and 2.5_1-helices.

Eq. F.4 was utilized to calculate the Ψ Ramachandran angles from the $\mathrm{AmIII^S}_3$ band frequencies of structures with complex or poorly defined solvation such as for turn-like conformations:

$$\upsilon_i = 1250\ \mathrm{cm^{-1}} - 54\mathrm{cm^{-1}}\sin(\Psi + 26) - 0.06\ \mathrm{cm^{-1}}/°C(T - T_0) \qquad \textbf{F.4}$$

where parameters T and T_0 are defined as in eq. F.2. Rearrangement of eq F.4 to solve for Ψ yields:

$$\Psi = \arcsin\left(\frac{\upsilon_{AmIII} - 1250 + 0.06(T - T_0)}{-54}\right) - 26 \qquad \textbf{F.5}$$

Eq. F.5 is recommended by Mikhonin et al.[64] for situations when the hydrogen bonding states of the peptide bond N–H groups are complex. This situation occurs, for example, in the case of

turn-like structures, where the peptide bonds can engage in both inter-amide and amide-water hydrogen bonding.

F.3 **Valid Range of Lorenzian Frequencies**

It is important to note that some Lorentzian band frequencies lie outside the valid range of our $AmIII_3^S$ ν_i -Ψ angle correlations. For example, Ψ < ~1200 and Ψ >~1310 values are invalid because they are just outside the valid sin Ψ domain frequency range, (Figure F.2).

To determine the valid limits of the $AmIII_3^S$ ν_i -Ψ correlation frequency range, we calculated the valid lower and upper limits of the $AmIII_3^S$ ν_i frequency in equations F.2 and F.4. In these equations, the maxima and minima of ν_i depend on the maximum and minimum values of the term -54 cm^{-1} sin(Ψ+26°), while all other terms are constant. We calculated ν_i for equations F.2 and F.4 to determine the maximum and minimum values of the quantity -54(cm^{-1}) sin(Ψ+26°).

For example, we determined the lower limit of the $AmIII_3^S$ ν_i frequency correlation by minimizing the term -54 cm^{-1}sin (Ψ+26) of eq F.2 (T=18°)

$$\upsilon_i = 1254\ \mathrm{cm}^{-1} - 54\ \mathrm{cm}^{-1}$$

and eq F.4 (T=18°)

$$\upsilon_i = 1249\ \mathrm{cm}^{-1} - 54\ \mathrm{cm}^{-1}$$

We calculated the lower limits of equations F.2 and F.4 to be 1200 cm^{-1} and 1195 cm^{-1}, respectively. We find that the ν_i minimum allowed Ψ values occur at Ψ=64°. Because of the resulting decrease in the $AmIII_3^S$ ν_i frequencies, additional regions of the Ψ angle distribution

become invalid solutions. Figure F.3 illustrates this point; the data for Ψ>40° are invalid, and are, thus, ignored.

By calculating ν_i for maximum values of the term -54cm^{-1} sin(Ψ+26), we determined that the upper limits of eqns F.2 and F.4 are ν_i=1308 cm^{-1} and ν_i=1303 cm^{-1}, respectively. The ν_i frequency maximum occurs at Ψ=-116°. Thus, ν_i frequency values >~1310 cm^{-1} are undefined.

An example of a Ψ angle outside the $AmIII_3^S$ -Ψ correlation range can be shown for an $AmIII_3^S$ frequency of 1190 cm^{-1} for a peptide with complicated hydrogen bonding (eq F.4):

$$\upsilon_i = 1250\ \text{cm}^{-1} - 54\text{cm}^{-1}\sin(\Psi + 26) - 0.06\ \text{cm}^{-1}/°C(T - T_0)$$

$$1190\ \text{cm}^{-1} = 1250\ \text{cm}^{-1} - 54\text{cm}^{-1}\sin(\Psi + 26) - 0.06\ \text{cm}^{-1}/°C(T - T_0)$$

Because $T-T_0$=18°:

$$1190\ \text{cm}^{-1} = 1250\ \text{cm}^{-1} - 54\text{cm}^{-1}\sin(\Psi + 26) - 1.08\ \text{cm}^{-1}$$

$$1190\ \text{cm}^{-1} = 1249\ \text{cm}^{-1} - 54\text{cm}^{-1}\sin(\Psi + 26)$$

$$-59\ \text{cm}^{-1} = -54\text{cm}^{-1}\sin(\Psi + 26)$$

$$1.09 = \sin(\Psi + 26)$$

$$\sin^{-1}(1.09) = \Psi + 26°$$

$\sin^{-1}(1.09)$is mathematically undefined!

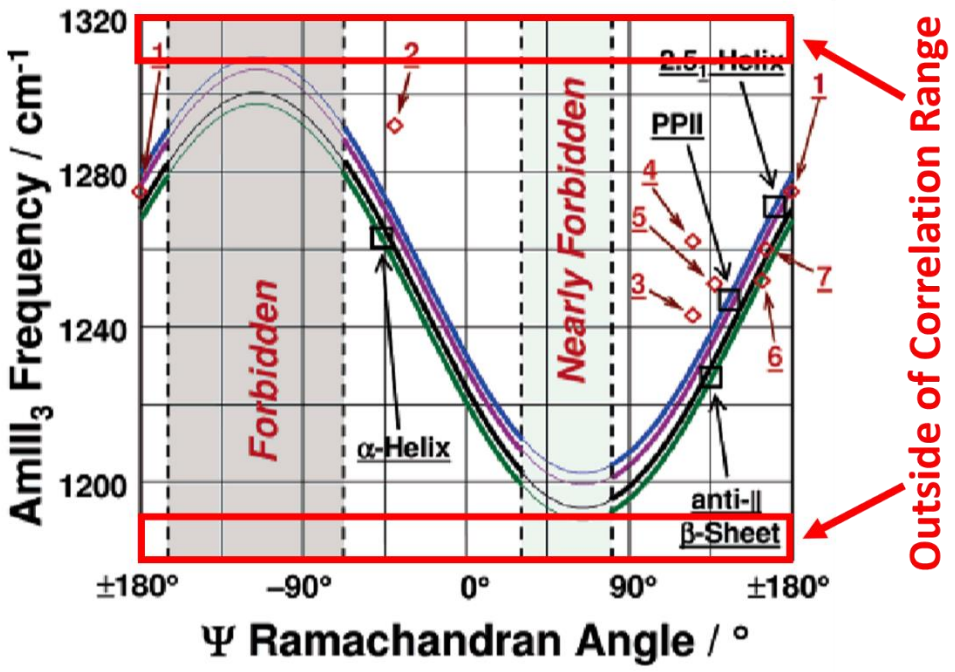

Figure F.2: Range of the $AmIII_3^S$ ν_i frequency-Ψ angle correlation. The red boxes highlight $AmIII_3^S$ frequencies that are outside of the valid correlation range.

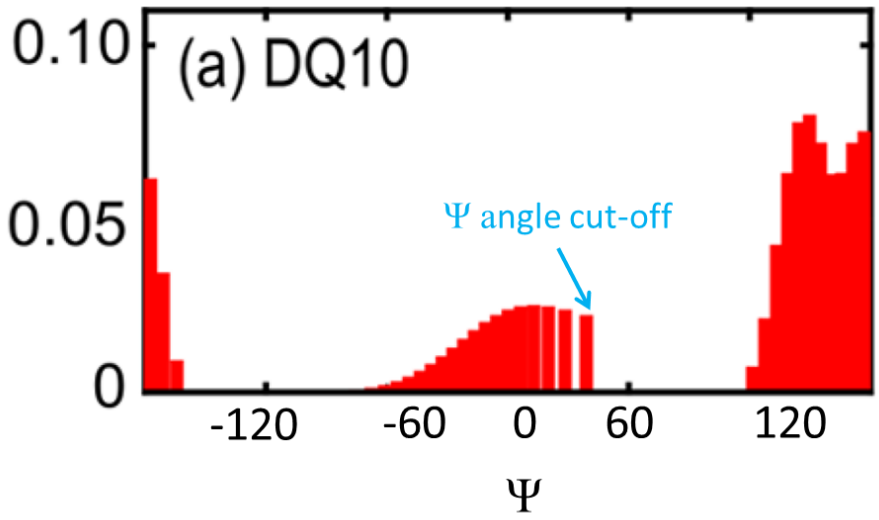

Figure F.3: Ψ angle distribution of the DQ10 peptide. The blue arrow indicates the exclusion of particular Ψ angles from the distribution. As discussed above, these excluded Ψ angles occur because the frequencies of the Lorentzian bands that model the Ψ angle distribution lie outside of the valid range of Ψ angles.

Our Ψ angle distributions will show gaps between data points as our Ψ angle distributions approach Ψ =64° (Figure F.4). To calculate the Ψ angle distributions, we use Mikhonin's equations to convert the center frequencies of our $AmIII_3^S$ frequency Lorentzian bands into Ψ angles.[64] Our Ψ angle distribution histograms consist of an array of discrete data points each represented by a histogram bar chosen to have a width of ~7°.

For adjacent Ψ angle distances of ≤ 7° the bars overlap. Because there is a sinusoidal $AmIII_3^S$ ν_i-Ψ correlation, the spacing between Ψ angle data points depends on the $AmIII_3^S$ frequency. At ~1230 cm^{-1} (Ψ = ~0°) $\delta\Psi/\delta\nu$ = ~1°/cm^{-1} resulting in a spacing between bars of ~5° and the bars overlap (Figure F.5). At ~1210 cm^{-1} (Ψ = ~30°) $\delta\Psi/\delta\nu$ = ~2°/cm^{-1} resulting in a spacing of ~10° and gaps appear between histogram data points as shown in Figure F.5.

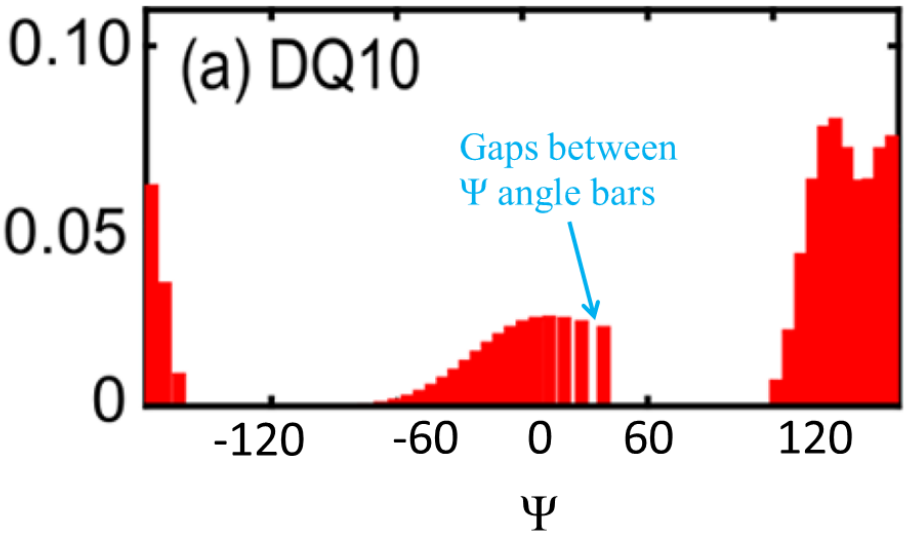

Figure F.4: Gaps between Ψ angle data points in the Ψ angle distribution of DQ10. These gaps are indicated by the blue arrow.

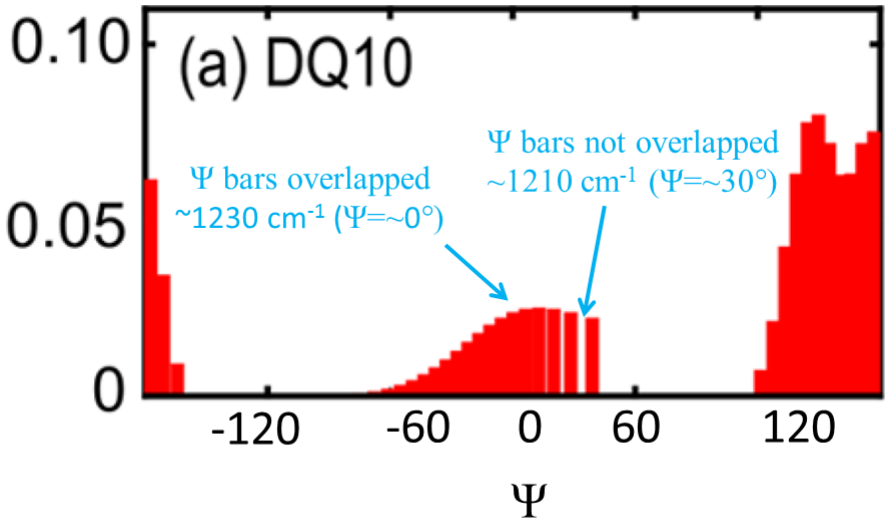

Figure F.5: Overlapped histogram bars of the Ψ angle distribution of DQ10. The blue arrows indicate regions where the spacing between data points show both overlapping and non-overlapping Ψ angle bars (see text for details).

12.0 Bibliography

(1) Orr, H. T.; Zoghbi, H. Y. *Annu. Rev. Neurosci.* **2007**, *30*, 575—621.

(2) Maciel, P.; Gaspar, C.; DeStefano, A. L.; Silveira, I.; Coutinho, P.; Radvany, J.; Dawson, D. M.; Sudarsky, L.; Guimarães, J.; Loureiro, J. E. L.; others. *Am. J. Hum. Genet.* **1995**, *57* (1), 54.

(3) Gatchel, J. R.; Zoghbi, H. Y. *Nat Rev Genet* **2005**, *6* (10), 743–755.

(4) Orr, H. T.; Chung, M.; Banfi, S.; Kwiatkowski Jr, T. J.; Servadio, A.; Beaudet, A. L.; McCall, A. E.; Duvick, L. A.; Ranum, L. P. W.; Zoghbi, H. Y. *Nat. Genet.* **1993**, *4* (3), 221.

(5) Myers, R. H. *NeuroRx* **2004**, *1* (2), 255–262.

(6) DiFiglia, M.; Sapp, E.; Chase, K. O.; Davies, S. W.; Bates, G. P.; Vonsattel, J. P.; Aronin, N. *Science (80-.).* **1997**, *277* (5334), 1990–1993.

(7) Becher, M. W.; Kotzuk, J. A.; Sharp, A. H.; Davies, S. W.; Bates, G. P.; Price, D. L.; Ross, C. A. *Neurobiol. Dis.* **1998**, *4* (6), 387–397.

(8) Ordway, J. M.; Tallaksen-Greene, S.; Gutekunst, C.-A.; Bernstein, E. M.; Cearley, J. A.; Wiener, H. W.; Dure IV, L. S.; Lindsey, R.; Hersch, S. M.; Jope, R. S.; others. *Cell* **1997**, *91* (6), 753–763.

(9) Marsh, J. L.; Walker, H.; Theisen, H.; Zhu, Y.-Z.; Fielder, T.; Purcell, J.; Thompson, L. M. *Hum. Mol. Genet.* **2000**, *9* (1), 13–25.

(10) Ikeda, H.; Yamaguchi, M.; Sugai, S.; Aze, Y.; Narumiya, S.; Kakizuka, A. *Nat. Genet.* **1996**, *13* (2), 196.

(11) Rigamonti, D.; Sipione, S.; Goffredo, D.; Zuccato, C.; Fossale, E.; Cattaneo, E. *J. Biol. Chem.* **2001**, *276* (18), 14545–14548.

(12) Gervais, F. G.; Singaraja, R.; Xanthoudakis, S.; Gutekunst, C.-A.; Leavitt, B. R.; Metzler, M.; Hackam, A. S.; Tam, J.; Vaillancourt, J. P.; Houtzager, V.; others. *Nat. Cell Biol.* **2002**, *4* (2), 95.

(13) Dragatsis, I.; Levine, M. S.; Zeitlin, S. *Nat. Genet.* **2000**, *26* (3), 300.

(14) Yang, W.; Dunlap, J. R.; Andrews, R. B.; Wetzel, R. *Hum. Mol. Genet.* **2002**, *11* (23), 2905–2917.

(15) Arrasate, M.; Mitra, S.; Schweitzer, E. S.; Segal, M. R.; Finkbeiner, S. *Nature* **2004**, *431* (7010), 805.

(16) Nagai, Y.; Inui, T.; Popiel, H. A.; Fujikake, N.; Hasegawa, K.; Urade, Y.; Goto, Y.; Naiki, H.; Toda, T. *Nat. Struct. Mol. Biol.* **2007**, *14* (4), 332–340.

(17) Trottier, Y.; Lutz, Y.; Stevanin, G.; Imbert, G.; Devys, D.; Cancel, G.; Saudou, F.; Weber, C.; David, G.; Tora, L.; Agid, Y.; Brice, A.; Mandel, J.-L. *Nature* **1995**, *378* (6555), 403–406.

(18) Bennett, M. J.; Huey-Tubman, K. E.; Herr, A. B.; West, A. P.; Ross, S. a; Bjorkman, P. J. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11634–11639.

(19) Li, P.; Huey-Tubman, K. E.; Gao, T.; Li, X.; West, A. P.; Bennett, M. J.; Bjorkman, P. J. *Nat. Struct. Mol. Biol.* **2007**, *14* (5), 381–387.

(20) Nelson, D. L.; Cox, M. M.; Lehninger, A. L. *Lehninger Principles of Biochemistry*, 7th ed.; Macmillan Learning, 2017.

(21) Horton, H. R.; Moran, L. A.; Scrimgeour, K. G.; Perry, M. D.; Rawn, J. D. *Principles of biochemistry*; Prentice Hall Upper Saddle River, NJ, 2006.

(22) Dunbrack, R. L.; Karplus, M. *Nat. Struct. Mol. Biol.* **1994**, *1* (5), 334–340.

(23) Shapovalov, M. V; Jr., R. L. D. *Structure* **2011**, *19* (6), 844–858.

(24) Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*; Macmillan, 1999.

(25) Estrada, E. *Bioinformatics* **2002**, *18* (5), 697–704.

(26) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44* (2), 97–179.

(27) Tonlolo, C.; Benedetti, E. *Trends Biochem. Sci.* **1991**, *16*, 350–353.

(28) Moradi, M.; Babin, V.; Roland, C.; Darden, T. A.; Sagui, C. *Proc. Natl. Acad. Sci.* **2009**, *106* (49), 20746–20751.

(29) Stapley, B. J.; Creamer, T. P. *Protein Sci.* **1999**, *8* (3), 587–595.

(30) Tompa, P. *Trends Biochem. Sci.* **2002**, *27* (10), 527–533.

(31) Chen, S.; Wetzel, R. *Protein Sci.* **2001**, *10* (4), 887–891.

(32) Perutz, M. F.; Johnson, T.; Suzuki, M.; Finch, J. T. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (12), 5355—5358.

(33) Chen, S.; Berthelier, V.; Yang, W.; Wetzel, R. *J. Mol. Biol.* **2001**, *311* (1), 173–182.

(34) Altschuler, E. L.; Hud, N. V; Mazrimas, J. a; Rupp, B. *J. Pept. Res.* **1997**, *50* (1), 73–75.

(35) Chen, S.; Berthelier, V.; Hamilton, J. B.; O'Nuallain, B.; Wetzel, R.; O'Nuallai, B.; Wetzel, R. *Biochemistry* **2002**, *41* (23), 7391–7399.

(36) Wetzel, R. *J. Mol. Biol.* **2012**, *421* (4–5), 466–490.

(37) Wang, X.; Vitalis, A.; Wyczalkowski, M. A.; Pappu, R. V. *Proteins* **2006**, *63* (2), 297–311.

(38) Vitalis, A.; Wang, X.; Pappu, R. V. *J. Mol. Biol.* **2008**, *384* (1), 279–297.

(39) Nakano, M.; Watanabe, H.; Rothstein, S. M.; Tanaka, S. *J. Phys. Chem. B* **2010**, *114* (20), 7056–7061.

(40) Bhattacharyya, A.; Thakur, A. K.; Chellgren, V. M.; Thiagarajan, G.; Williams, A. D.; Chellgren, B. W.; Creamer, T. P.; Wetzel, R. *J. Mol. Biol.* **2006**, *355* (3), 524–535.

(41) Klein, F. a C.; Pastore, A.; Masino, L.; Zeder-Lutz, G.; Nierengarten, H.; Oulad-Abdelghani, M.; Altschuh, D.; Mandel, J.-L.; Trottier, Y. *J. Mol. Biol.* **2007**, *371* (1), 235–244.

(42) Vitalis, A.; Lyle, N.; Pappu, R. V. *Biophys. J.* **2009**, *97* (1), 303–311.

(43) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (45), 16764–16769.

(44) Walters, R. H.; Murphy, R. M. *J. Mol. Biol.* **2009**, *393* (4), 978–992.

(45) Vitalis, A.; Wang, X.; Pappu, R. V. *Biophys. J.* **2007**, *93* (6), 1923–1937.

(46) Punihaole, D.; Workman, R. J.; Hong, Z.; Madura, J. D.; Asher, S. A. *J. Phys. Chem. B* **2016**, *120* (12), 3012–3026.

(47) Chen, S.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci.* **2002**, *99* (18), 11884–11889.

(48) Thakur, A. K.; Jayaraman, M.; Mishra, R.; Thakur, M.; Chellgren, V. M.; L Byeon, I.-J.; Anjum, D. H.; Kodali, R.; Creamer, T. P.; Conway, J. F.; M Gronenborn, A.; Wetzel, R. *Nat Struct Mol Biol* **2009**, *16* (4), 380–389.

(49) Sikorski, P.; Atkins, E. *Biomacromolecules* **2005**, *6* (1), 425–432.

(50) Schneider, R.; Schumacher, M. C.; Mueller, H.; Nand, D.; Klaukien, V.; Heise, H.; Riedel, D.; Wolf, G.; Behrmann, E.; Raunser, S.; Seidel, R.; Engelhard, M.; Baldus, M. *J. Mol. Biol.* **2011**, *412* (1), 121–136.

(51) Perutz, M. F.; Finch, J. T.; Berriman, J.; Lesk, A. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (8), 5591–5595.

(52) Hoop, C. L.; Lin, H.-K.; Kar, K.; Magyarfalvi, G.; Lamley, J. M.; Boatz, J. C.; Mandal, A.; Lewandowski, J. R.; Wetzel, R.; van der Wel, P. C. A. *Proc. Natl. Acad. Sci.* **2016**, *113* (6), 1546–1551.

(53) Hoop, C. L.; Lin, H.-K.; Kar, K.; Hou, Z.; Poirier, M. A.; Wetzel, R.; van der Wel, P. C. A. *Biochemistry* **2014**, *53* (42), 6653–6666.

(54) Lin, H.-K.; Boatz, J. C.; Krabbendam, I. E.; Kodali, R.; Hou, Z.; Wetzel, R.; Dolga, A. M.; Poirier, M. A.; van der Wel, P. C. A. *Nat. Commun.* **2017**, *8*.

(55) Buchanan, L. E.; Carr, J. K.; Fluitt, A. M.; Hoganson, A. J.; Moran, S. D.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T. *Proc. Natl. Acad. Sci.* **2014**, *111* (16), 5796–5801.

(56) Moran, S. D.; Zanni, M. T. *J. Phys. Chem. Lett.* **2014**, *5* (11), 1984–1993.

(57) Xiong, K.; Punihaole, D.; Asher, S. A. *Biochemistry* **2012**, *51* (29), 5822–5830.

(58) Kar, K.; Jayaraman, M.; Sahoo, B.; Kodali, R.; Wetzel, R. *Nat Struct Mol Biol* **2011**, *18* (3), 328–336.

(59) Thakur, A. K.; Wetzel, R. *Proc. Natl. Acad. Sci.* **2002**, *99* (26), 17014–17019.

(60) Berthelier, V.; Hamilton, J. B.; Chen, S.; Wetzel, R. *Anal. Biochem.* **2001**, *295* (2), 227–236.

(61) Vitalis, A.; Pappu, R. V. *Biophys. Chem.* **2011**, *159* (1), 14–23.

(62) Oladepo, S. A.; Xiong, K.; Hong, Z.; Asher, S. A.; Handen, J.; Lednev, I. K. *Chem. Rev.* **2012**, *112* (5), 2604–2628.

(63) Jakubek, R. S.; Handen, J.; White, S. E.; Asher, S. A.; Lednev, I. K. *Trends Anal. Chem.* **2018**, *103*, 223–229.

(64) Mikhonin, A. V; Bykov, S. V; Myshakina, N. S.; Asher, S. A. *J. Phys. Chem. B* **2006**, *110* (4), 1928–1943.

(65) Asher, S. A.; Ianoul, A.; Mix, G.; Boyden, M. N.; Karnoup, A.; Diem, M.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2001**, *123* (47), 11775–11781.

(66) Asher, S. A.; Larkin, P. J.; Teraoka, J. *Biochemistry* **1991**, *30* (24), 5944–5954.

(67) Larkin, P. J.; Gustafson, W. G.; Asher, S. A. *J. Chem. Phys.* **1991**, *94* (8), 5324–5330.

(68) Chi, Z.; Asher, S. a. *J. Phys. Chem. B* **1998**, *102* (47), 9595–9602.

(69) Ahmed, Z.; Myshakina, N. S.; Asher, S. a. *J. Phys. Chem. B* **2009**, *113* (32), 11252–11259.

(70) Hong, Z.; Wert, J.; Asher, S. A. *J. Phys. Chem. B* **2013**, *117* (24), 7145–7156.

(71) Punihaole, D.; Hong, Z.; Jakubek, R. S.; Dahlburg, E. M.; Geib, S.; Asher, S. A. *J. Phys. Chem. B* **2015**, *119* (41), 13039–13051.

(72) Punihaole, D.; Jakubek, R. S.; Dahlburg, E. M.; Hong, Z.; Myshakina, N. S.; Geib, S.; Asher, S. A. *J. Phys. Chem. B* **2015**, *119* (10), 3931–3939.

(73) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. *J. Am. Chem. Soc.* **1999**, *121* (35), 8074–8086.

(74) Asher, S. a; Mikhonin, A. V; Bykov, S. *J. Am. Chem. Soc.* **2004**, *126* (27), 8433–8440.

(75) Mikhonin, A. V; Asher, S. a. *J. Am. Chem. Soc.* **2006**, *128* (42), 13789–13795.

(76) Ma, L.; Hong, Z.; Sharma, B.; Asher, S. *J. Phys. Chem. B* **2012**, *116* (3), 1134–1142.

(77) Long, D. A. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*; John Wiley and Sons Ltd.: Chichester, 2002.

(78) Xiong, K.; Ma, L.; Asher, S. A. *Biophys. Chem.* **2012**, *162*, 1–5.

(79) Hong, Z.; Damodaran, K.; Asher, S. A. *J. Phys. Chem. B* **2014**, *118* (36), 10565–10575.

(80) Xiong, K.; Asher, S. A. *J. Phys. Chem. B* **2012**, *116* (24), 7102–7112.

(81) Punihaole, D.; Jakubek, R. S.; Workman, R. J.; Asher, S. A. *J. Phys. Chem. Lett.* **2018**, *9* (8), 1944–1950.

(82) Punihaole, D.; Jakubek, R. S.; Workman, R. J.; Marbella, L. E.; Campbell, P.; Madura, J. D.; Asher, S. A. *J. Phys. Chem. B* **2017**, *121* (24), 5953–5967.

(83) Asher, S. A.; Ludwig, M.; Johnson, C. R. *J. Am. Chem. Soc.* **1986**, *108* (12), 3186–3197.

(84) Ludwig, M.; Asher, S. A. *J. Am. Chem. Soc.* **1988**, *110* (4), 1005–1011.

(85) Chi, Z.; Asher, S. A. *Biochemistry* **1999**, *38* (26), 8196–8203.

(86) Chen, X. G.; Asher, S. A.; Schweitzer-Stenner, R.; Mirkin, N. G.; Krimm, S. *J. Am. Chem. Soc.* **1995**, *117* (10), 2884–2895.

(87) Albrecht, A. C.; Vleck, V. *J. Chem. Phys.* **1961**, *34* (5), 1476–1484.

(88) Hong, Z.; Asher, S. A. *Appl. Spectrosc.* **2015**, *69* (1), 75–83.

(89) Mayne, L. C.; Hudson, B. *J. Phys. Chem.* **1991**, *95* (8), 2962–2967.

(90) Myshakina, N. S.; Ahmed, Z.; Asher, S. A. *J. Phys. Chem. B* **2008**, *112* (38), 11873–11877.

(91) Markham, L. M.; Hudson, B. S. *J. Phys. Chem.* **1996**, *100* (7), 2731–2737.

(92) Hudson, B. S.; Markham, L. M. *J. Raman Spectrosc.* **1998**, *29* (6), 489–500.

(93) Wang, Y.; Purrello, R.; Ceorgiou, S.; Spiro, T. G.; Georgiou, S.; Spiro, T. G. *J. Am. Chem. Soc.* **1991**, *113* (17), 6368–6377.

(94) Siebrand, W.; Zgierski, M. Z. *J. Chem. Phys.* **1979**, *71* (9), 3561–3569.

(95) Asher, S. A. *Annu. Rev. Phys. Chem.* **1988**, *39* (1), 537–588.

(96) Asher, S. A.; Chi, Z.; Li, P. *J. Raman Spectrosc.* **1998**, *29* (10–11), 927–931.

(97) Murphy, R. M.; Walters, R. H.; Tobelmann, M. D.; Bernacki, J. P. In *Non-fibrillar Amyloidogenic Protein Assemblies-Common Cytotoxins Underlying Degenerative Diseases*; Springer, 2012; pp 337–375.

(98) Chellgren, B. W.; Miller, A.-F.; Creamer, T. P. *J. Mol. Biol.* **2006**, *361* (2), 362–371.

(99) Vitalis, A.; Lyle, N.; Pappu, R. V. *Biophys. J.* **2009**, *97* (1), 303–311.

(100) Singh, V. R.; Lapidus, L. J. *J. Phys. Chem. B* **2008**, *112* (42), 13172–13176.

(101) Ladurner, A. G.; Itzhaki, L. S.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci.* **1998**, *95* (15), 8473–8478.

(102) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. *Proc. Natl. Acad. Sci.* **2013**, *110* (17), 6817–6822.

(103) Bykov, S.; Lednev, I.; Ianoul, A.; Mikhonin, A.; Munro, C.; Asher, S. A. *Appl. Spectrosc.* **2005**, *59* (12), 1541–1552.

(104) Piotto, M.; Saudek, V.; Sklená\vr, V. *J. Biomol. NMR* **1992**, *2* (6), 661–665.

(105) Stejskal, E. O.; Tanner, J. E. *J. Chem. Phys.* **1965**, *42* (1), 288–292.

(106) Molecular Operating Environment. Chemical Computing group Inc.: Montreal, QC 2016.

(107) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.

(108) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (1), 33–38.

(109) Huang, J.; MacKerell, A. D. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145.

(110) MacKerell Jr, A. D. *J. Comput. Chem.* **2004**, *25* (13), 1584–1604.

(111) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(112) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101* (5), 4177–4189.

(113) Asciutto, E. K.; Mikhonin, A. V; Asher, S. A.; Madura, J. D. **2008**, 2046–2050.

(114) Asciutto, E. K.; General, I. J.; Xiong, K.; Xiong, K.; Asher, S. a; Madura, J. D. *Biophys. J.* **2010**, *98* (2), 186–196.

(115) Xiong, K.; Asciutto, E. K.; Madura, J. D.; Asher, S. A. *Biochemistry* **2009**, *48* (45), 10818–10826.

(116) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100* (2), 20603.

(117) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12562–12566.

(118) Mikhonin, A. V.; Ahmed, Z.; Ianoul, A.; Asher, S. a. *J. Phys. Chem. B* **2004**, *108* (49), 19020–19028.

(119) Mikhonin, A. V; Myshakina, N. S.; Bykov, S. V; Asher, S. a. *J. Am. Chem. Soc.* **2005**, *127* (21), 7712–7720.

(120) Walsh, P. S.; Blodgett, K. N.; McBurney, C.; Gellman, S. H.; Zwier, T. S. *Angew. Chemie Int. Ed.* **2016**, *55* (47), 14618–14622.

(121) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17* (1), 75–86.

(122) Gekko, K.; Ohmae, E.; Kameyama, K.; Takagi, T. *Biochim. Biophys. Acta (BBA)-Protein Struct. Mol. Enzymol.* **1998**, *1387* (1), 195–205.

(123) Ensing, B.; De Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39* (2), 73–81.

(124) Vymetal, J.; Vondrasek, J. *J. Phys. Chem. B* **2010**, *114* (16), 5632–5642.

(125) Laio, A.; Gervasio, F. L. *Reports Prog. Phys.* **2008**, *71* (12), 126601.

(126) Henin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. *J. Chem. Theory Comput.* **2009**, *6* (1), 35–47.

(127) Gaborek, T. J.; Chipot, C.; Madura, J. D. *Biophys. J.* **2012**, *103* (12), 2513–2520.

(128) Kar, K.; Hoop, C. L.; Drombosky, K. W.; Baker, M. a; Kodali, R.; Arduini, I.; van der Wel, P. C. a; Horne, W. S.; Wetzel, R. *J. Mol. Biol.* **2013**, *425* (7), 1183–1197.

(129) Milner-White, E. J.; Poet, R. *Biochem. J.* **1986**, *240* (1), 289–292.

(130) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell Jr, A. D. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257.

(131) Tiffany, M. L.; Krimm, S. *Biopolymers* **1968**, *6* (9), 1379–1382.

(132) Tifany, M. L.; Krimm, S. *Biopolymers* **1972**, *11* (11), 2309–2316.

(133) Krimm, S.; Mark, J. E. *Proc. Natl. Acad. Sci.* **1968**, *60* (4), 1122–1129.

(134) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, *90* (4), L36--L38.

(135) Feng, L. Molecular Dynamics Predicts The Solution Conformations of Poly-L-lysine in Salt Solutions, University of Pittsburgh, 2014.

(136) Moradi, M.; Babin, V.; Roland, C.; Sagui, C. *PLoS Comput Biol* **2012**, *8* (4), e1002501.

(137) Miettinen, M. S.; Knecht, V.; Monticelli, L.; Ignatova, Z. *J. Phys. Chem. B* **2012**, *116* (34), 10259–10265.

(138) Darnell, G. D.; Derryberry, J.; Kurutz, J. W.; Meredith, S. C. *Biophys. J.* **2009**, *97* (8), 2295–2305.

(139) Bhattacharyya, A. M.; Thakur, A. K.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15400–15405.

(140) Heck, B. S.; Doll, F.; Hauser, K. *Biophys. Chem.* **2014**, *185*, 47–57.

(141) Mangiarini, L.; Sathasivam, K.; Seller, M.; Cozens, B.; Harper, A.; Hetherington, C.; Lawton, M.; Trottier, Y.; Lehrach, H.; Davies, S. W.; others. *Cell* **1996**, *87* (3), 493–506.

(142) Scherzinger, E.; Lurz, R.; Turmaine, M.; Mangiarini, L.; Hollenbach, B.; Hasenbank, R.; Bates, G. P.; Davies, S. W.; Lehrach, H.; Wanker, E. E. *Cell* **1997**, *90* (3), 549–558.

(143) Chellgren, B. W.; Miller, A.-F.; Creamer, T. P. *J. Mol. Biol.* **2006**, *361* (2), 362–371.

(144) Rhys, N. H.; Soper, A. K.; Dougan, L. *J. Phys. Chem. B* **2012**, *116* (45), 13308–13319.

(145) Natalello, A.; Frana, A. M.; Relini, A.; Apicella, A.; Invernizzi, G.; Casari, C.; Gliozzi, A.; Doglia, S. M.; Tortora, P.; Regonesi, M. E. *PLoS One* **2011**, *6* (4), e18789.

(146) Deng, H.; Callender, R. *Meth. Enzym. Suppl. C* **1999**, *308*, 176–201.

(147) Badger, R. M.; Bauer, S. H. *J. Chem. Phys.* **1937**, *5* (11), 839–851.

(148) Levinson, N. M.; Fried, S. D.; Boxer, S. G. *J. Phys. Chem. B* **2012**, *116* (35), 10470–10476.

(149) Schneider, S. H.; Kratochvil, H. T.; Zanni, M. T.; Boxer, S. G. *J. Phys. Chem. B* **2017**, *121* (10), 2331–2338.

(150) Choi, J.-H.; Oh, K.-I.; Lee, H.; Lee, C.; Cho, M. *J. Chem. Phys.* **2008**, *128* (13), 134506.

(151) Pazos, I. M.; Ghosh, A.; Tucker, M. J.; Gai, F. *Angew. Chemie* **2014**, *126* (24), 6194–6198.

(152) Riddle Jr, F. L.; Fowkes, F. M. *J. Am. Chem. Soc.* **1990**, *112* (9), 3259–3264.

(153) Varfolomeev, M. A.; Rakipov, I. T.; Solomonov, B. N. *Int. J. Thermophys.* **2013**, *34* (4), 710–724.

(154) Cutmore, E. A.; Hallam, H. E. *Spectrochim. Acta Part A Mol. Spectrosc.* **1969**, *25* (11), 1767–1784.

(155) Eaton, G.; Symons, M. C. R.; Rastogi, P. P. *J. Chem. Soc. Faraday Trans. 1 Phys. Chem. Condens. Phases* **1989**, *85* (10), 3257–3271.

(156) Mayer, U.; Gutmann, V.; Gerger, W. *Monatshefte f{ü}r Chemie/Chemical Mon.* **1975**, *106* (6), 1235–1257.

(157) Sivanandam, V. N.; Jayaraman, M.; Hoop, C. L.; Kodali, R.; Wetzel, R.; van der Wel, P. C. A. *J. Am. Chem. Soc.* **2011**, *133* (12), 4558–4566.

(158) Moore, W. H.; Krimm, S. *Proc. Natl. Acad. Sci.* **1975**, *72* (12), 4933–4935.

(159) Moore, W. H.; Krimm, S. *Biopolymers* **1976**, *15* (12), 2439–2464.

(160) Moore, W. H.; Krimm, S. *Biopolymers* **1976**, *15* (12), 2465–2483.

(161) Miyazawa, T. *J. Chem. Phys.* **1960**, *32* (6), 1647–1652.

(162) Klotz, I. M.; Franzen, J. S. *J. Am. Chem. Soc.* **1962**, *84* (18), 3461–3466.

(163) Davies, M.; Thomas, D. K. *J. Phys. Chem.* **1956**, *60* (6), 767–770.

(164) Spencer, J. N.; Garrett, R. C.; Mayer, F. J.; Merkle, J. E.; Powell, C. R.; Tran, M. T.; Berger, S. K. *Can. J. Chem.* **1980**, *58* (13), 1372–1375.

(165) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(166) Ross, C. A.; Poirier, M. A. *Nat. Rev. Mol. cell Biol.* **2005**, *6* (11), 891–898.

(167) Penney, J. B.; Vonsattel, J.-P.; Macdonald, M. E.; Gusella, J. F.; Myers, R. H. *Ann. Neurol.* **1997**, *41* (5), 689–692.

(168) Sugaya, K.; Matsubara, S.; Kagamihara, Y.; Kawata, A.; Hayashi, H. *PLoS One* **2007**, *2* (7), e635.

(169) Sugaya, K.; Matsubara, S. *Mol. Neurodegener.* **2009**, *4* (1), 29.

(170) Asher, S. A.; Mikhonin, A. V; Bykov, S. *J. Am. Chem. Soc.* **2004**, *126* (27), 8433–8440.

(171) Mayne, L. C.; Hudson, B. *J. Phys. Chem.* **1991**, *95* (8), 2962–2967.

(172) Heller, E. J.; Sundberg, R.; Tannor, D. *J. Phys. Chem.* **1982**, *86* (10), 1822–1833.

(173) Barth, A. *Biochim. Biophys. Acta (BBA)-Bioenergetics* **2007**, *1767* (9), 1073–1101.

(174) Valenti, L. E.; Paci, M. B.; De Pauli, C. P.; Giacomelli, C. E. *Anal. Biochem.* **2011**, *410* (1), 118–123.

(175) Averill, B. A.; Eldredge, P. *General Chemistry: Principles, Patterns, and Applications*; 2012.

(176) Robinson, R. E.; Taylor, R. C. *Spectrochim. Acta* **1962**, *18*, 1093–1097.

(177) Klemperer, W.; Pimentel, G. C. *J. Chem. Phys.* **1954**, *22* (8), 1399–1402.

(178) Fuson, N.; Josien, M.-L.; Jones, E. A.; Lawson, J. R. *J. Chem. Phys.* **1952**, *20* (10), 1627–1634.

(179) Dudik, J. M.; Johnson, C. R.; Asher, S. A. *J. Chem. Phys.* **1985**, *82* (4).

(180) Burra, G.; Thakur, A. K. *Data Br.* **2015**, *5*, 1066–1071.

(181) Burra, G.; Thakur, A. K. *Anal. Biochem.* **2016**, *494*, 23–30.

(182) Sreerama, N.; Woody, R. W. *Proteins Struct. Funct. Bioinforma.* **1999**, *36* (4), 400–406.

(183) Adzhubei, A. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1993**, *229* (2), 472–493.

(184) Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. *Proteins Struct. Funct. Bioinforma.* **2004**, *55* (3), 502–507.

(185) Drozdov, A. N.; Grossfield, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2004**, *126* (8), 2574–2581.

(186) Fleming, P. J.; Fitzkee, N. C.; Mezei, M.; Srinivasan, R.; Rose, G. D. *Protein Sci.* **2005**, *14* (1), 111–118.

(187) Makarov, A. A.; Lobachov, V. M.; Adzhubei, I. A.; Esipova, N. G. *FEBS Lett.* **1992**, *306* (1), 63–65.

(188) Eker, F.; Cao, X.; Nafie, L.; Huang, Q.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2003**, *107* (1), 358–365.

(189) Burra, G.; Thakur, A. K. *Pept. Sci.* e24094.

(190) Lajoie, P.; Snapp, E. L. *PLoS One* **2010**, *5* (12), e15245.

(191) Liu, K.-Y.; Shyu, Y.-C.; Barbaro, B. A.; Lin, Y.-T.; Chern, Y.; Thompson, L. M.; James Shen, C.-K.; Marsh, J. L. *Hum. Mol. Genet.* **2014**, *24* (6), 1602–1616.

(192) Chen, S.; Berthelier, V.; Yang, W.; Wetzel, R. *J. Mol. Biol.* **2001**, *311* (1), 173–182.

(193) Jakubek, R. S.; White, S. E.; Asher, S. A. *J. Phys. Chem. B.* **2019**, *123* (8), 1749–1763.

(194) Dudik, J. M.; Johnson, C. R.; Asher, S. A. *J. Phys. Chem.* **1985**, *89* (18), 3805–3814.

(195) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109* (14), 6714–6721.

(196) Hovmöller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2002**, *58* (5), 768–776.

(197) Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. *Biochemistry* **2005**, *44* (28), 9691–9702.

(198) Woody, R. W. In *Circular Dichroism and the Conformational Analysis of Biomolecules*; Fasman, G. D., Ed.; Plenum Press, New York, 1996; pp 25–67.

(199) Darnell, G.; Orgel, J. P. R. O.; Pahl, R.; Meredith, S. C. *J. Mol. Biol.* **2007**, *374* (3), 688–704.

(200) Michalik, A.; Van Broeckhoven, C. *Hum. Mol. Genet.* **2003**, *12* (suppl_2), R173--R186.

(201) Masino, L.; Kelly, G.; Leonard, K.; Trottier, Y.; Pastore, A. *FEBS Lett.* **2002**, *513* (2–3), 267–272.

(202) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31* (6), 1647–1651.

(203) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4* (2), 171–180.

(204) Jakubek, R. S.; Workman, R. J.; White, S. E.; Asher, S. A. *J. Phys. Chem. B* **2019**, *123* (19), 4193–4203.

(205) Danielsson, J.; Jarvet, J.; Damberg, P.; Gräslund, A. *Magn. Reson. Chem.* **2002**, *40* (13).

(206) Chan, T. C.; Chan, M. L. *J. Chem. Soc. Faraday Trans.* **1992**, *88* (16), 2371–2374.

(207) Chan, T. C.; Tang, W. K. *J. Chem. Phys.* **2013**, *138* (22), 224503.

(208) Lu, J. G.; Kong, R.; Chan, T. C. *J. Chem. Phys.* **1999**, *110* (6), 3003–3008.

(209) Hayduk, W.; Buckley, W. D. *Chem. Eng. Sci.* **1972**, *27* (11), 1997–2003.

(210) Schneider, R.; Schumacher, M. C.; Mueller, H.; Nand, D.; Klaukien, V.; Heise, H.;

Riedel, D.; Wolf, G.; Behrmann, E.; Raunser, S.; Seidel, R.; Engelhard, M.; Baldus, M.; et al. *J. Mol. Biol.* **2011**, *412* (1), 121–136.

(211) Isas, J. M.; Langen, R.; Siemer, A. B. *Biochemistry* **2015**, *54* (25), 3942–3949.

(212) Wu, D. H.; Chen, A. D.; Johnson, C. S. *J. Magn. Reson. Ser. A* **1995**, *115* (2), 260–264.

(213) Cotts, R. M.; Hoch, M. J. R.; Sun, T.; Markert, J. T. *J. Magn. Reson.* **1989**, *83* (2), 252–266.

(214) Tanner, J. E. *J. Chem. Phys.* **1970**, *52* (5), 2523–2526.

(215) Chi, Z.; Chen, X. G.; Holtz, J. S. W.; Asher, S. A. *Biochemistry* **1998**, *37* (9), 2854–2864.

(216) Englander, S. W.; Downer, N. W.; Teitelbaum, H. *Annu. Rev. Biochem.* **1972**, *41* (1), 903–924.

(217) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36* (8), 1627–1639.

(218) Sreerama, N.; Woody, R. W. *Biochemistry* **1994**, *33* (33), 10022–10025.

(219) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (14), 9190–9195.

(220) Harris, D. C. *Quantitative Chemical Analysis*, 7th ed.; W.H. Freeman and Company: New York, 2007.

(221) A. Bundi, K. W. *Biopolymers* **1979**, *18* (2), 285–297.

(222) Zamyatnin, A. A. *Annu. Rev. Biophys. Bioeng.* **1984**, *13* (1), 145–165.