

**THE KINETICS AND THERMODYNAMICS OF POLYGLUTAMINE  
AGGREGATION**

by

Elizabeth Christine Landrum

Bachelor of Arts, Cornell University, 2008

Submitted to the Graduate Faculty of  
The School of Medicine in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Elizabeth Christine Landrum

It was defended on

December 6, 2013

and approved by

Seth Horne, Ph.D., Department of Chemistry

Patrick H. Thibodeau, Ph.D., Department of Cell Biology and Physiology

Daniel M. Zuckerman, Ph.D., Department of Computational & Systems Biology

Dissertation Advisor: Ronald B. Wetzel, Ph.D., Department of Structural Biology

Copyright © by Elizabeth Landrum

2013

# THE KINETICS AND THERMODYNAMICS OF POLYGLUTAMINE AGGREGATION

Elizabeth Christine Landrum, PhD

University of Pittsburgh, 2013

Protein misfolding as a result of polyglutamine (polyQ) repeat expansion plays a crucial role in the development a class of neurodegenerative disorders. These disorders are characterized by expansions in unique proteins, but share one common factor: polyQ repeat expansions beyond a pathological threshold, usually near 35 Glns. PolyQ containing peptides and proteins aggregate into amyloid-like fibrils, in both disease states and *in vitro*. While recent studies of simple polyQ aggregation have led to the development of a mechanism describing the fibril formation pathway, the kinetic and thermodynamic roles of even small modifications to the amino acid sequence are unclear. The goal of this work was to gain more specific information regarding the biophysical roles of terminal charge and Gln repeat length in the aggregation mechanism of simple polyQ peptides. We found that increases in both repeat length and attraction between terminal flanking sequences led to enhancements in fibril stability and spontaneous aggregation ability. In general, mutations in polyQ sequences that enhance their ability to assume a  $\beta$ -hairpin structure (the proposed building block of polyQ fibrils), and to associate with each other facilitate fibrillization through a nucleated polymerization pathway with a monomeric nucleus. The data discussed here have made significant contributions to our understanding of the basic aggregation properties of polyQ peptides, and have implications that can be directly translated to polyQ disease pathogenesis.

## TABLE OF CONTENTS

<b>ABBREVIATIONS</b> .....	<b>XIV</b>
<b>CONTRIBUTOR NAMES</b> .....	<b>XV</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>XVI</b>
<b>1.0 INTRODUCTION</b> .....	<b>1</b>
<b>1.1 AMYLOID FIBRILS</b> .....	<b>1</b>
<b>1.1.1 Amyloid Structure and Properties</b> .....	<b>1</b>
<b>1.1.2 Amyloid Aggregation</b> .....	<b>3</b>
<b>1.2 POLYGLUTAMINE REPEAT DISORDERS</b> .....	<b>7</b>
<b>1.2.1 General features</b> .....	<b>7</b>
<b>1.2.2 Role of Context in PolyQ Containing Proteins</b> .....	<b>11</b>
<b>1.2.3 Physiological Features of PolyQ Disorders</b> .....	<b>12</b>
<b>1.3 AGGREGATION OF POLYGLUTAMINE CONTAINING PEPTIDES</b> ...	<b>14</b>
<b>1.3.1 <i>In Vitro</i> PolyQ Aggregation</b> .....	<b>14</b>
<b>1.3.2 Structural Steps of PolyQ Fibril Formation</b> .....	<b>15</b>
<b>1.3.2.1 The PolyQ Monomer Ensemble</b> .....	<b>15</b>
<b>1.3.2.2 The Toxic Monomer Hypothesis</b> .....	<b>18</b>
<b>1.3.2.3 PolyQ Oligomers and Non-fibrillar aggregates</b> .....	<b>19</b>
<b>1.3.2.4 PolyQ Amyloid</b> .....	<b>20</b>

1.3.2.5	<b>β-turn inducing mutations.....</b>	<b>22</b>
1.3.2.6	<b>Summary.....</b>	<b>23</b>
1.3.3	<b>Flanking Sequences in PolyQ Aggregation .....</b>	<b>24</b>
1.3.3.1	<b>Solubilizing charges are required to study simple polyQ aggregation.....</b>	<b>24</b>
1.3.3.2	<b>Disease relevant flanking sequences can impact aggregation mechanism.....</b>	<b>26</b>
1.3.4	<b>PolyQ Demonstrates Repeat Length Dependent Aggregation .....</b>	<b>28</b>
1.3.5	<b>Mathematical Models for PolyQ Aggregation .....</b>	<b>30</b>
1.3.5.1	<b>A Nucleated Polymerization Model of Aggregation .....</b>	<b>30</b>
1.3.5.2	<b>Alternate Models of Aggregation.....</b>	<b>34</b>
1.3.6	<b>Fundamental Questions Being Addressed.....</b>	<b>35</b>
1.3.6.1	<b>Flanking sequences in polyQ peptide aggregation.....</b>	<b>35</b>
1.3.6.2	<b>Repeat length dependence .....</b>	<b>35</b>
1.4	<b>PRELIMINARY STUDIES .....</b>	<b>36</b>
1.4.1	<b>Peptide Design.....</b>	<b>36</b>
1.4.1.1	<b>Optimization of peptide length .....</b>	<b>36</b>
2.0	<b>EXPERIMENTAL METHODS.....</b>	<b>40</b>
2.1	<b>PEPTIDE PREPARATION AND REACTION PROTOCOLS .....</b>	<b>40</b>
2.1.1	<b>Peptide Synthesis and Purification .....</b>	<b>40</b>
2.1.2	<b>Peptide Disaggregation and Reaction Preparation .....</b>	<b>41</b>
2.1.3	<b>Peptide mass confirmation by mass spectrometry .....</b>	<b>42</b>
2.1.4	<b>Peptide concentration determination.....</b>	<b>42</b>

2.1.5	Dissociation reactions .....	43
2.2	ANALYSIS OF AGGREGATION KINETICS .....	44
2.2.1	Sedimentation Assay.....	44
2.2.2	Thioflavin T Binding .....	44
2.2.3	Dynamic light scattering .....	45
2.2.4	Seeded elongation reactions.....	46
2.2.5	Growing ends reaction .....	47
2.2.6	Aggregation and nucleation kinetics.....	49
3.0	METHOD DEVELOPMENT: UNDERSTANDING THE GROWING ENDS TITRATION EXPERIMENT.....	51
3.1	OVERVIEW.....	51
3.2	RESULTS .....	54
3.2.1	Biotinyl PolyQ Elongation of PolyQ Fibrils .....	54
3.2.2	Time dependent elongation with biotinyl-polyQ .....	58
3.2.3	Growing ends titration accurately reports available growing ends.....	60
3.2.3.1	A first attempt: varying the seed mass.....	61
3.2.3.2	Sonication generates samples with variable amounts of elongation sites	62
3.3	DISCUSSION.....	66
4.0	TERMINAL CHARGE STUDIES OF POLYQ PEPTIDES.....	70
4.1	OVERVIEW.....	70
4.2	RESULTS .....	72
4.2.1	Impact of Terminal Charges on Kinetics and Morphology.....	72

4.2.2	Fibril Stability .....	74
4.2.3	Fibril Elongation.....	78
4.2.4	Application of Nucleated Polymerization Model.....	80
4.2.4.1	Nucleus size determination.....	80
4.2.4.2	Efficiency of spontaneous aggregation.....	83
4.2.4.3	Nucleus stability calculations yield confusing results .....	84
4.3	DISCUSSION.....	86
5.0	REPEAT LENGTH DEPENDENT EFFECTS ON POLYQ AGGREGATION.	91
5.1	OVERVIEW.....	91
5.2	RESULTS .....	92
5.2.1	Fibril stability.....	92
5.2.2	Spontaneous aggregation and nucleation .....	94
5.2.3	Elongation rates and nucleation equilibrium constants.....	101
5.3	DISCUSSION.....	103
6.0	CONCLUSIONS .....	105
6.1	THESIS CONTRIBUTIONS.....	105
6.1.1	Use of growing ends titration assay.....	105
6.1.2	Terminal charge contributions.....	106
6.1.3	PolyQ repeat length.....	108
6.1.4	Summary .....	109
6.2	FUTURE DIRECTIONS.....	110
6.2.1	Does peptide net charge or solubility have a greater impact on mechanism? .....	110

6.2.2	Can we generate a monomeric nucleus from low repeat length peptides?	110
6.2.3	Big picture .....	111
APPENDIX A	.....	112
BIBLIOGRAPHY	.....	120

## LIST OF TABLES

Table 1-1. PolyQ Repeat Disorders .....	8
Table 1-2. PolyQ Protein Features.....	11
Table 1-3. Hairpin encouraging mutations affect n* .....	22
Table 3-1. Sonication impacts on growing ends.....	64
Table 4-1. Terminal charge peptides and charge characteristics.....	71
Table 4-2. Terminal charge thermodynamic values.....	77
Table 5-1. Repeat length peptides and features.....	92

## LIST OF FIGURES

Figure 1-1. $\beta$ -arcade structures. ....	2
Figure 1-2. Amyloid formation occurs <i>via</i> a complex pathway. ....	6
Figure 1-3. Age of disease onset is dependent on PolyQ repeat length.....	10
Figure 1-4. Structure of PolyQ $\beta$ -strands.....	20
Figure 1-5. Aggregation kinetics of hairpin encouraging peptides [19].....	23
Figure 1-6. Thermodynamic model of polyQ aggregation. ....	31
Figure 1-7. Positive and negative effects on hairpin closure due to terminal charge interactions. .....	37
Figure 1-8. Aggregation of $K_2Q_{10}P^DGGQ_{10}K_2$ and $K_2Q_9P^DGGQ_9K_2$ . ....	39
Figure 3-1. Growing ends assay.....	52
Figure 3-2. Biotinyl polyQ seeding reactions. ....	55
Figure 3-3. EM micrographs of biotinyl-polyQ fibrils. ....	57
Figure 3-4. Time dependence of biotinyl-polyQ incorporation.....	59
Figure 3-5. Growing ends titration data from reactions with variable amounts of seed.....	61
Figure 3-6. Seeded elongation and growing ends data (14 krpm centrifugation).....	63
Figure 3-7. Variable seed size growing ends titration (ultracentrifugation). ....	64
Figure 3-8. Seeded elongation and growing ends titration with variably sized seeds (filter trap). .....	65
Figure 3-9. Proposed mechanism for growing ends titration assay. ....	67

Figure 4-1. Aggregation kinetics of all terminal charge peptides.....	73
Figure 4-2. EM images of end stage aggregates. ....	74
Figure 4-3. Association and dissociation reactions meet at the $C_r$ . ....	75
Figure 4-4. Free energy of elongation vs. repulsion. ....	76
Figure 4-5. Elongation rate constant vs. repulsion. ....	78
Figure 4-6. Elongation rate constant vs. fibril stability. ....	80
Figure 4-7. Concentration dependent kinetics, $\text{time}^2$ and log-log plots of $K_2Q_{26}K_2$ . ....	81
Figure 4-8. Complex parameter vs. repulsion.....	83
Figure 4-9. $K_{n^*}$ vs. repulsion.....	85
Figure 5-1. $\Delta G_{\text{elong}}$ versus repeat length. ....	93
Figure 5-2. $Q_{42}$ kinetics, $\text{time}^2$ and log-log plots. ....	94
Figure 5-3. A fit of y-intercept versus repeat length for peptides with monomeric nuclei.....	95
Figure 5-4. Calculation of monomeric y-intercepts for short polyQ peptides.....	96
Figure 5-5. New fit of y-intercept vs. repeat length for peptides with monomeric nuclei.....	99
Figure 5-6. $\Delta y\text{-int/residue}$ versus repeat length. ....	100
Figure 5-7. $k_+$ versus repeat length. ....	102
Figure 5-8. $K_{n^*}$ versus repeat length. ....	102
Figure A-1. $KQ_{26}K_3$ kinetics, $\text{time}^2$ , and log-log plot.....	112
Figure A-2. $Q_{26}K$ kinetics, $\text{time}^2$ , and log-log plot. ....	113
Figure A-3. $KQ_{26}D$ kinetics, $\text{time}^2$ , and log-log plot. ....	113
Figure A-4. $K_2Q_9P^D GQ_9K_2$ kinetics, $\text{time}^2$ , and log-log plot. ....	113
Figure A-5. $KQ_9P^D GQ_9K_3$ kinetics, $\text{time}^2$ , and log-log plot.....	114
Figure A-6. $K Q_9P^D GQ_9K$ kinetics, $\text{time}^2$ , and log-log plot. ....	114

Figure A-7. $Q_9P^D GQ_9K$ kinetics, $\text{time}^2$ , and log-log plot. ....	114
Figure A-8. $KQ_9P^D GQ_9D$ kinetics, $\text{time}^2$ , and log-log plot. ....	115
Figure A-9. $K_2Q_9P^D GQ_9D_2$ kinetics, $\text{time}^2$ , and log-log plot. ....	115
Figure A-10. Reactions of $K_2Q_9P^D GQ_9K_2$ at various pH, salt concentrations, and temperatures. .....	116
Figure A-11. $K_2Q_{27}K_2$ kinetics, $\text{time}^2$ , and log-log plot. ....	117
Figure A-12. $K_2Q_{29}K_2$ kinetics, $\text{time}^2$ , and log-log plot. ....	118
Figure A-13. $K_2Q_{30}K_2$ kinetics, $\text{time}^2$ , and log-log plot. ....	118
Figure A-14. $K_2Q_{34}K_2$ kinetics, $\text{time}^2$ , and log-log plot. ....	118
Figure A-15. $K_2Q_{37}K_2$ kinetics, $\text{time}^2$ , and log-log plot. ....	119

## ABBREVIATIONS

HD	Huntington's disease
Htt	Huntingtin
NT	N-terminal
polyQ	polyglutamine
CD	circular diocroism
EM	electron microscopy
AFM	atomic force microscopy
ThT	thioflavin T
PPII	Polyproline helix type-II
SAXS	Small-angle x-ray scattering
FRET	Förster Resonance Energy transfer
FTIR	Fourier-transform infra-red spectroscopy
ssNMR	solid state nuclear magnetic resonance
TFA	2, 2, 2-Trifluoroacetic acid
HFIP	1, 1,1,3,3,3-Hexafluoro-2-propanol
RP-HPLC	Reverse-phase High performance liquid chromatography
UV	Ultraviolet
DLS	Dynamic light scattering
MD	Molecular dynamics
SDS	sodium dodecyl sulfata

PBS            phosphate buffered saline

### **CONTRIBUTOR NAMES**

MJ            Murali Jayaraman

KK            Karunakar Kar

CH            Cody Hoop

PVW          Patrick van der Wel

BS            Bankanidhi Sahoo

RK            Ravindra Kodali

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my dissertation advisor, Ron Wetzels, for his mentorship, patience and inspiration these past few years. I would also like to thank my committee along with the rest of the MBSB faculty—without their help and insight, this work would not be what it is today.

My fellow MBSB and Pitt School of Medicine students have served as friends, mentors, troubleshooters, and teachers during my Ph.D.. I would especially like to thank Cody Hoop, who has never failed to listen and lend a helping hand, whether with problems of a scientific or personal nature. My labmates, Bankanidhi Sahoo, in particular, have served as great teachers during my time in the Wetzels lab. If not for their willingness to sit down and discuss science at the drop of a hat, I would not be near as qualified as I am today to write this thesis. Also, Ravindra Kodali, who helped with the preparation and acquisition of all the EM data shown in this thesis.

The staff within the Graduate Studies Office have been invaluable during my Ph.D. Without regular help from Susanna Godwin and Jennifer Walker, navigating the joint university MBSB program would have been beyond challenging.

Finally, I would like to thank my close friends and family who have supported me these past five and a half years and served as my cheerleaders, even if they had no idea what I was doing.

## 1.0 INTRODUCTION

### 1.1 AMYLOID FIBRILS

#### 1.1.1 Amyloid Structure and Properties

The ability to polymerize and form amyloid fibrils is a characteristic that seems to be inherent to the nature of polypeptide chains [1-5]. Misfolding and aggregation of specific cellular proteins is an important causative factor in a number of serious diseases, over 20 of which are characterized by the formation of amyloid fibrils [3, 6]. In spite of their implication in disease, formation of amyloid fibrils *in vivo* has also been suggested to sometimes play a functional or protective role [7-9].

Amyloid fibrils can be identified by their fibrillar morphology (as observed by electron microscopy (EM) or atomic force microscopy (AFM)), enhancement of thioflavin-T (ThT) fluorescence upon binding, and their birefringence in polarized microscopy as enhanced by Congo red staining [3, 10]. They also share a set of common structural attributes, in spite of the fact that each disease is characterized by aggregation of a particular protein. The most notable feature of amyloid fibrils is that they are made up primarily of cross- $\beta$  sheet, as identified *via* the cross- $\beta$  x-ray diffraction pattern of aligned fibrils [3]. In cross- $\beta$  sheet,  $\beta$ -strands lie orthogonal to the fibril axis and assemble into  $\beta$ -sheet *via* backbone H-bonds that run parallel to the fibril axis.

Sheet-sheet packing is stabilized through side-chain interactions perpendicular to the fibril axis [11]. Most fibrils consist of parallel, in-register  $\beta$ -sheet, though some fibrils have been found to consist of pseudo-in register parallel strands, and only rarely (polyQ fibrils) are anti-parallel sheets observed [12].

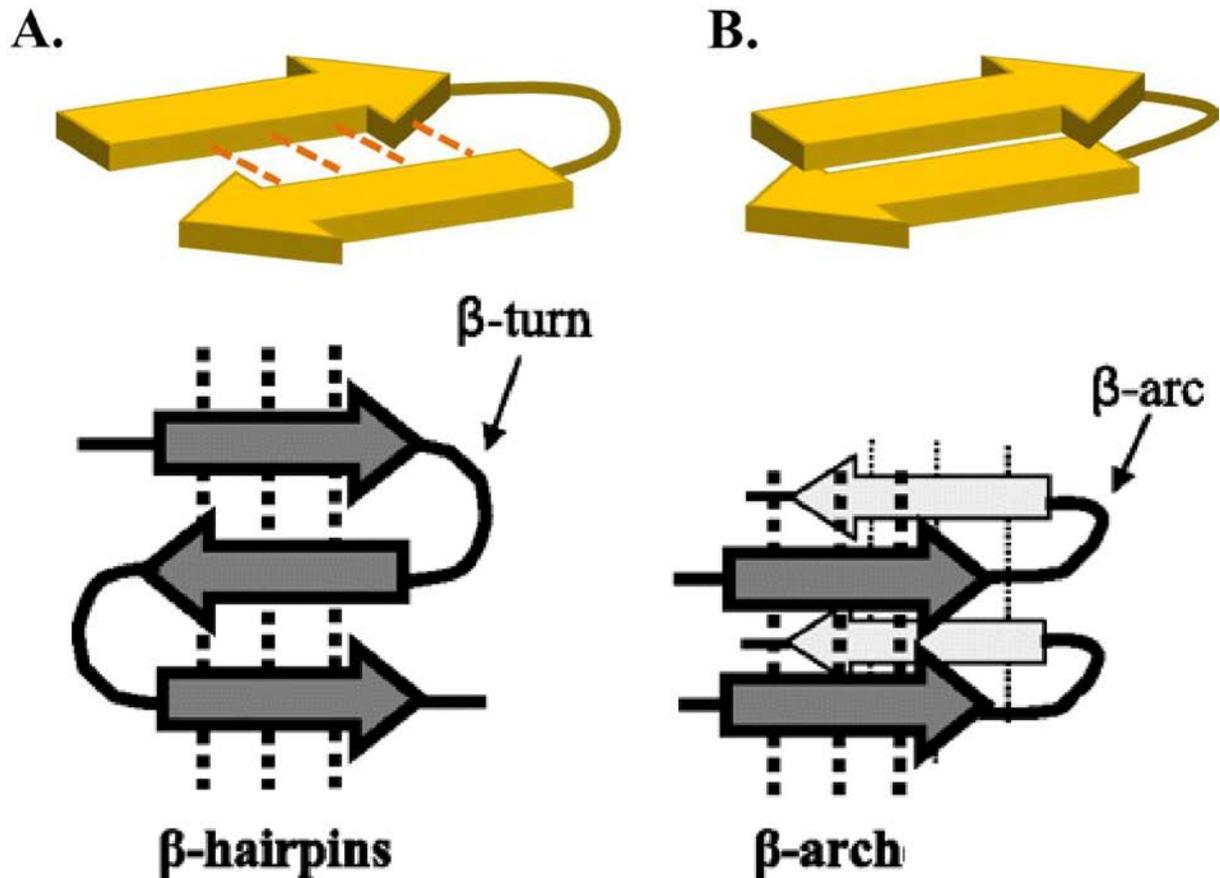


Figure 1-1.  $\beta$ -arcade structures.

A  $\beta$ -hairpin (A), consists of two  $\beta$ -strands of at least 7 or 8 residues each linked by a turn or loop. This structure is stabilized predominantly through main-chain—main chain hydrogen bonding (orange dashed lines). The  $\beta$ -arch (B), is similarly made up of two  $\beta$ -strands linked by a turn. In this case, the structure is primarily formed by intramolecular side chain interactions and stabilized as additional molecules bind, establishing intermolecular main chain hydrogen bonds.

Except for fibrils formed by very short peptides, amyloid fibrils nearly always comprise of a series of  $\beta$ -strands linked by turns and/or loops [13-15]. The  $\beta$ -arch is commonly observed in amyloid, and has been identified as the building block of a number of fibrils, including amyloid- $\beta$  ( $A\beta$ ) and yeast prion proteins [14]. In this structure, intrachain  $\beta$ -strands stack on top of each other, linked by a large turn or loop (**Figure 1-1 B**). The  $\beta$ -arch is stabilized by intramolecular side chain interactions and intermolecular main-chain hydrogen bonds that form as additional molecules bind. The antiparallel or parallel nature of the growing  $\beta$ -sheet is determined by the orientation of elongating molecules.  $\beta$ -hairpins (**Figure 1-1 A**) consist of a pair of  $\beta$ -strands aligned side-by-side and linked by a turn or loop, forming a two strand antiparallel sheet stabilized by intramolecular main-chain hydrogen bonds. [16, 17].  $\beta$ -hairpins are uncommon in amyloids described to date, but have been proposed as building blocks of the antiparallel  $\beta$ -sheet containing polyQ fibrils (described in more detail in section 1.3.2) [18-20].

### 1.1.2 Amyloid Aggregation

In the most general sense, formation of amyloid fibrils is a protein misfolding reaction. More specifically, however, there are a number of possible mechanisms by which a single protein or peptide can become a part the highly-structured complex that is an amyloid fibril (**Figure 1-2**).

It is generally accepted that amyloid formation entails a number of features of a nucleated growth polymerization mechanism, wherein a structural nucleus is formed, followed by rapid association of additional molecules resulting in formation of the amyloid fibril. The particular nature of the nucleus (monomeric vs. multimeric, its exact structure, etc.) and mechanism of elongation (monomer addition vs. association of oligomers or complexes) vary from peptide to peptide.

While a primary nucleus serves as the catalyst to promote amyloid formation, fibrils are also capable of secondary nucleation *via* fragmentation of the fibril or branching [21, 22]. When this occurs, new, previously inaccessible growing ends are presented to the reaction. Kinetic contributions from secondary nucleation events make it nearly impossible to extract information about primary nucleation from the aggregation kinetics [22-24]. Because of this, systems which do not undergo secondary nucleation as a part of their polymerization mechanisms provide ideal opportunities to glean information about the primary nucleation phase of the aggregation mechanism.

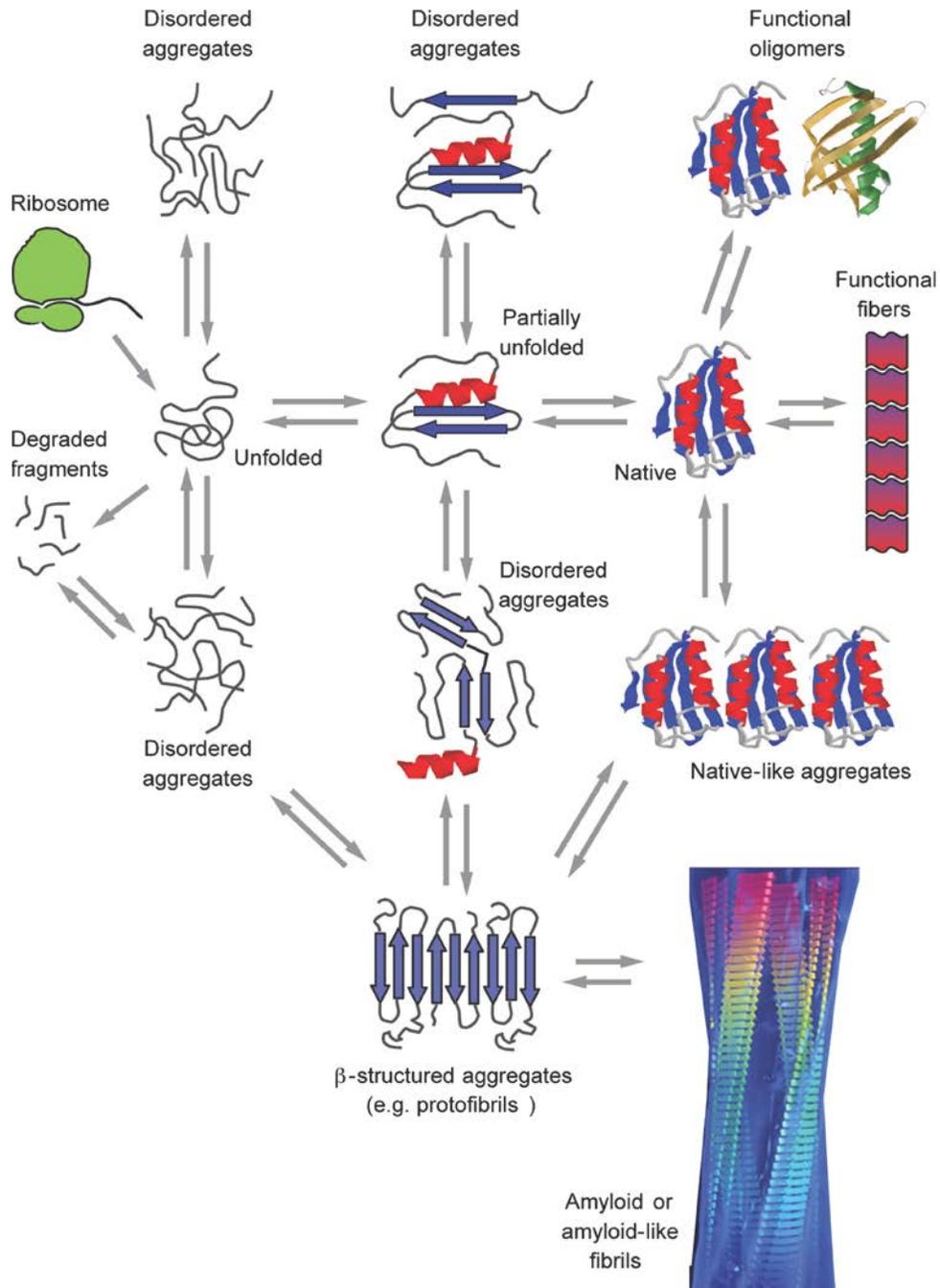
Once nucleation occurs, the polymerization process proceeds by addition of molecules (whether by monomer addition or otherwise) to the growing ends of the fibril, and continues until an equilibrium is reached. At this point, the monomer concentration stabilizes at a value referred to as the critical concentration ( $C_r$ ).  $C_r$  can be regarded as an indication of the stability of the fibril, and corresponds to the inverse association constant for fibril growth ( $K_a^{-1}$ ). Thus, the  $C_r$  can be used identically to  $K_a^{-1}$  to calculate the free energy of elongation,  $\Delta G_{\text{elong}} = -RT \ln(1/C_r)$ . This approach has been validated with Ala scanning mutations in A $\beta$ (1-40) [25]. The  $C_r$  values of a series of Ala mutations were used to determine  $\Delta G_{\text{elong}}$  values, which, in turn, were used to calculate the  $\Delta\Delta G_{\text{elong}}$  values for each mutation by comparing to the wild-type  $\Delta G_{\text{elong}}$  value ( $\Delta\Delta G_{\text{elong}} = \Delta G_{\text{elongWT}} - \Delta G_{\text{elongMut}}$ ). When matched to analogous mutations at adjacent positions in parallel  $\beta$ -sheet in the globular GB1 protein, these  $\Delta\Delta G$  values showed good agreement, indicating that the use of  $C_r$  values to determine  $\Delta G_{\text{elong}}$  is a robust approach to measure fibril stability in comparative studies.

In a number of cases, oligomeric species have been identified as intermediates to amyloid formation [21, 26-28]. The increased interest in these oligomeric species has been driven by their

proposed role as the toxic species in amyloidosis [29, 30]. Some of these oligomers appear to be on-pathway intermediates, such that when they are isolated and used as seeds, amyloid formation is accelerated [27, 31], but this is not always the case [32]. In the A $\beta$  system, although many non-fibrillar assemblies have been identified, most oligomers can be labeled as one of two key types: soluble oligomers that are low in molecular weight and disorganized in structure, or chain-like structures with diameters of 2-5 nm that are rich in  $\beta$ -sheet (observed *via* EM and AFM) [28, 33]. It is this second species of oligomer, referred to as a protofibril, that appears to be an on-pathway intermediate of amyloid formation (**Figure 1-2**). The smaller, soluble oligomers, have been observed in *in vivo* studies and are believed to play a toxic role in disease pathogenesis [29].

Elongation of amyloid fibrils is a more complex process than the simple association of incoming molecules to the growing end of the fibril. Elongation can occur *via* monomer addition, or by assembly of oligomers or protofibrils [27, 34]. Even in the case of monomer addition, the newly bound monomer must undergo conformational changes in order to become stably incorporated and present a new growing end to the system. Thus, fibril elongation is believed to be a multi-step process, wherein the elongating monomer adds *via* a “dock and lock” mechanism [35, 36]. The “docking” phase occurs when a disordered monomer encounters and binds to the growing end of a fibril (or nucleus). It is followed by one or more “locking” steps, in which the molecule adding to the fibril folds, assuming the  $\beta$ -sheet rich structure that is capable of supporting subsequent growth. Surface plasmon resonance (SPR) studies of the A $\beta$  peptide revealed that it elongates *via* a 3 step mechanism, with two kinetically defined “locking” steps [35]. The six individual elongation and dissociation rates measured during this study could be propagated into an overall equilibrium constant which corresponds very well to the

independently determined  $K_a$  value for  $A\beta$  determined from  $C_r$  [37] (further validating the use of the  $C_r$  value as a measure of fibril stability).



 Chiti F, Dobson CM. 2006. Annu. Rev. Biochem. 75:333–66

**Figure 1-2. Amyloid formation occurs *via* a complex pathway.**

From [30].

In summary, amyloid fibril formation is a complex process, during which a single molecule undergoes a series of conformational changes as it proceeds toward its ultimate fibrillar state [30]. As with any protein folding reaction, an unfolded molecule samples a number of states as it folds, some of which may involve association with other molecules into oligomers or higher order aggregates (**Figure 1-2**). In some cases, the formation of these complexes is on-pathway, with oligomers capable of acting as fibril nuclei or elongation material. In other cases these oligomers are off-pathway, forming structures not directly related to the fibril and whose only role is to tie up molecules and at least transiently prevent their involvement in the aggregation reaction. In the simplest case, disordered monomer is capable of undergoing a conformational change to nucleate amyloid formation and elongate *via* monomer addition. This process is discussed in more detail in Section 1.3.5.1.

## 1.2 POLYGLUTAMINE REPEAT DISORDERS

### 1.2.1 General features

Trinucleotide repeat expansions have been implicated in more than 16 neurodegenerative diseases [38-40]. CAG repeat disorders are a disease family characterized by the expansion of CAG nucleotides. When repeats of this codon occur in coding regions of the genome, the result is a polyQ expansion, which has been linked to 10 of the more than 20 amyloidoses (**Table 1**) [3, 4, 30, 41, 42]. While the focus of this work is on the polyQ containing polypeptides resulting from expressed CAG repeat containing DNA [43], some diseases are associated with trinucleotide expansions in non-coding regions of the genome. One disorder, HD-like-2, was

originally believed to fall into this group, since the CAG repeat expansion occurs in an anti-sense portion of DNA. However, the anti-sense strand is transcribed and translated, resulting in a protein product which has been implicated in a HD-like-2 [42, 44]. Because of the common features shared among these diseases, it is generally accepted that the polyQ containing protein is the key player in disease [41]. More specifically, neuronal toxicity has been proposed to be a result of a gain of function caused by expansion of the polyQ tract [40, 45, 46].

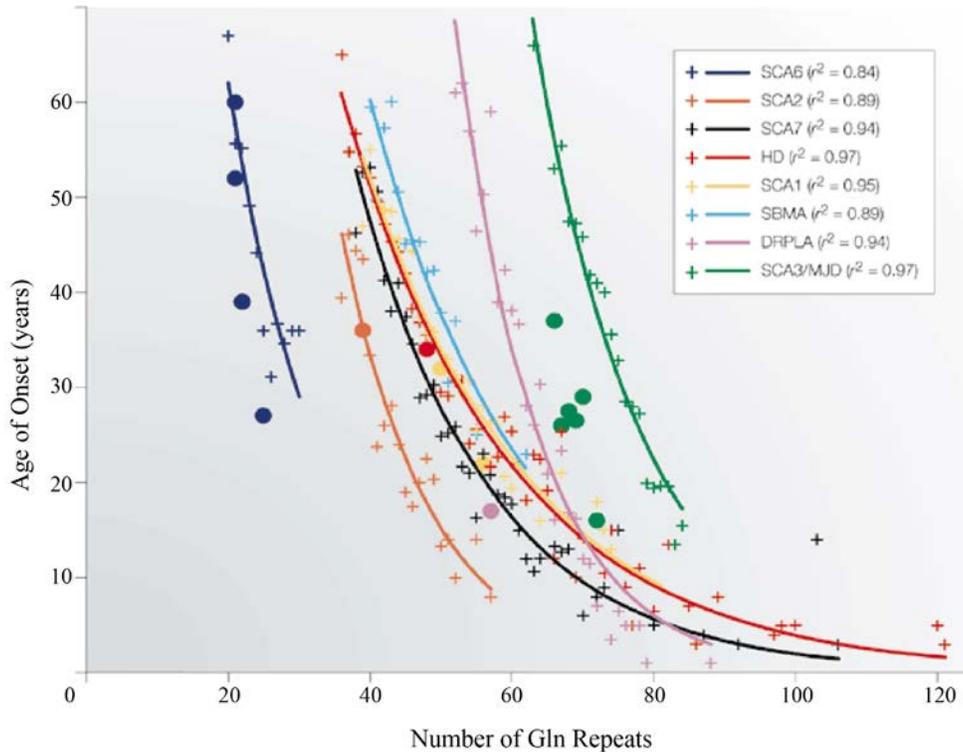
**Table 1-1. PolyQ Repeat Disorders**

<b>Disease</b>	<b>Protein</b>	<b>Normal Repeat Length</b>	<b>Expanded Repeat Length</b>	<b>Clinical Features</b>
<b>Huntington's Disease (HD)</b>	Huntingtin (exon 1)	6-34	36-121	Chorea, dystonia, cognitive deficits, psychiatric problems
<b>Spinocerebellar ataxia 1 (SCA1)</b>	Ataxin 1	6-44	39-82	Ataxia, slurred speech, spasticity, cognitive impairments
<b>SCA2</b>	Ataxin 2	15-24	32-200	Ataxia, polyneuropathy, decreased reflexes, infantile variant with retinopathy
<b>SCA3</b>	Ataxin 3	13-36	61-84	Ataxia, parkinsonism, spasticity
<b>SCA6</b>	CACNA1 <sub>A</sub>	4-19	10-33	Ataxia, dysarthria, nystagmus, tremors
<b>SCA7</b>	Ataxin 7	4-35	37-306	Ataxia, blindness, cardiac failure in infantile form
<b>SCA17</b>	TATA-binding protein (TBP)	25-42	47-63	Ataxia, cognitive decline, seizures, and psychiatric problems
<b>Spinobulbar muscular atrophy (SBMA: Kennedy Disease)</b>	Androgen receptor	9-36	38-62	Motor weakness, swallowing, gynecomastia, decreased fertility
<b>Dentatorubropallidoluysian atrophy (DRPLA)</b>	Atrophin	7-34	49-88	Ataxia, seizures, choreoathetosis, dementia
<b>HD-like 2</b>	*	7-27	40-60	abnormal movements, dementia, and psychiatric syndromes.

Modified from [40], supplemented with [44]. \* Since this disease protein results from translation from an anti-sense DNA strand, it does not correspond to a normally expressed protein. The sense strand, which codes for a CAG repeat, however, produces Junctophilin-3 (exon 2A).

The critical Gln repeat length threshold for amyloid in disease tends to lie in the mid-thirties, with most diseases only manifesting symptoms when repeat lengths reach 35-45 Gln residues [41]. In some cases, such as spinocerebellar ataxia 6 (SCA6), the physiological repeat length threshold is much shorter, at approximately 16 to 20 Gln residues [47, 48]. This difference in repeat length threshold is possibly due to differences in sequence context, an idea that will be discussed in more detail below (1.2.2). As the Gln repeat length surpasses the threshold at which disease symptoms manifest, age of onset decreases. In the extreme case, very long polyQ repeats lead to highly aggressive juvenile forms of the disease (**Figure 1-3**) [41, 49, 50].

The 10 proteins implicated in polyQ disorders range in size from 339 to 3144 residues (normal length). Their functionality is either unknown, or widely varies, as does their localization within the cell— the TATA-binding protein (implicated in SCA17) is localized to the nucleus and plays a role in transcription. CACNA1<sub>A</sub> (implicated in SCA6) is the  $\alpha_{1A}$  subunit of a voltage-gated calcium channel, and has only been observed to function in the cytoplasm. The androgen receptor containing the polyQ expansion associated with SMBA is a ligand activated transcription factor, found in the cytoplasm of neurons and the nuclei of non-neuronal tissue.



**Figure 1-3. Age of disease onset is dependent on PolyQ repeat length.**

The average age of onset for a given repeat length is indicated by a (+). An exponential decay best fit is given by smooth lines. SCA6 clearly presents at the shortest repeat lengths, with most other diseases first manifesting at Gln repeat counts near 40. Filled circles indicate patients homozygous for a polyQ expansion, while crosses are heterozygotes. Modified from [50].

The sequence context and location of the polyQ tract vary widely from protein to protein. These and other characteristics of polyQ proteins are listed in **Table 1-2**. Of particular interest is the fact that, in several of these proteins, polyQ repeat expansions are flanked by Pro rich regions. It has been suggested that these Pro containing regions exist to modulate the effects of the polyQ region by inducing formation of PPII helices [51, 52]. The effects of poly-Pro regions have been well studied *in vitro* and are discussed in more detail later in this chapter (1.3.3.2).

**Table 1-2. PolyQ Protein Features<sup>1</sup>.**

<b>Protein</b>	<b>Protein Size (number of amino acids)</b>	<b>Q Expansion Start Location</b>	<b>N-terminal Charge/Features</b>	<b>C-terminal Charge/Features</b>
<b>Huntingtin</b>	3144	18	+1	-65/Poly-Pro adjacent to polyQ region
<b>Ataxin 1</b>	816	197	+3/Pro, Gly, Ala, Ser make up >40%, His interrupt polyQ	+1/Pro and Ser rich
<b>Ataxin 2</b>	1313	166	+18/Pro content >20%, Gly, Ala, Ser rich	+30/poly-Pro adjacent to polyQ
<b>Ataxin 3</b>	359	292	-21/Gln, Leu make up >20%	-2/Ser, Leu make up >20%
<b>CACNA1<sub>A</sub></b>	2410	2328	+23/poly-Pro adjacent to polyQ	+9/Ala, Arg, Gly, Pro, Ser make up >50%
<b>Ataxin 7</b>	893	30	+1/Ala rich	+47/poly-Pro adjacent to polyQ, rich in Ser and Pro
<b>TBP</b>	339	58	-2/Gln and Pro rich	+15
<b>Androgen receptor</b>	918	58	+3/polyLeu adjacent to polyQ	-14/short polyQ repeats
<b>Atrophin</b>	1184	474	+5/Pro, Ser rich, His repeats adjacent to polyQ	+4/Ala, Pro, Ser rich, His repeats adjacent to polyQ

### 1.2.2 Role of Context in PolyQ Containing Proteins

The differences in repeat length dependence of disease onset across polyQ disorders (i.e. SCA6) suggest the specific protein containing the polyQ expansion could play an integral role in disease manifestation. Context dependence of these disorders is further supported by characteristic neuronal degeneration patterns unique to each disease [53, 54]. Modifications to flanking sequences have been found to play meaningful roles in toxicity and disease progression [55]. For

---

<sup>1</sup> All sequence information was acquired *via* UniProt for the human sequence of the particular protein. A sequence was called rich in a given amino acid if it was present in an amount  $\geq 10\%$ .

example, post translational modifications to flanking sequences in SCA1 and huntingtin exon 1 have notable effects on toxicity in disease models [56-59], as well as on aggregation propensity *in vitro* [60]. These data indicate that toxicity and disease pathogenesis is controlled by polyQ containing proteins as a whole, rather than solely through the polyQ containing region.

### 1.2.3 Physiological Features of PolyQ Disorders

Despite the fact that polyQ diseases are caused by Gln expansions within entirely unique proteins, they each demonstrate similar pathogenic presentations. All polyQ disorders result in neuronal dysfunction and localized neuronal degeneration, in spite of ubiquitous expression of these proteins [43]. The unique neurodegeneration patterns within each disease are likely partly due to expression levels and cellular response to the presence of these proteins and their misfolding—lower expression levels and the presence of cellular machinery capable of regulating misfolded proteins have been found to correspond to decreases in neurodegeneration [41]—in addition to their sequence context, as mentioned above (1.2.2).

Neurological aggregates are observed in all polyQ disorders. These aggregates have been described as amyloid, or amyloid-like [61-63], and *in vitro* studies of polyQ containing peptides have been found to generate amyloid fibrils that have similar properties to the aggregates found *in vivo* [62, 64]. Additionally, aggregate formation has been demonstrated to have a similar repeat dependence as disease onset (discussed below, 1.3.4) [18, 65]. Because of the common features in disease pathology across polyQ disorders, as well as the similar repeat length dependence in aggregate formation and pathogenesis, these aggregates (or a structure related to them) have been proposed to be the toxic species in polyQ disease.

Neuronal nuclear aggregates of the polyQ containing protein have been identified in each of the polyQ diseases [41, 53, 66]. These nuclear aggregates have been found to be ubiquitin positive, confirming that they are misfolded protein (ubiquitinylation is a post-translational modification that signals degradation of misfolded proteins [67]). As a common feature of all polyQ disorders, these nuclear inclusions were hypothesized to cause toxicity through nuclear dysfunction [62, 68]. This hypothesis was further supported by cell studies which introduced chemically synthesized polyQ peptide aggregates into the nuclei of neuronal cell cultures, demonstrating that the presence of polyQ aggregates in the nucleus leads to toxicity, while introduction of aggregates to the cytoplasm of these cells had a minimal impact on cell viability [69]. In conflict with this hypothesis, nuclear inclusions containing misfolded polyQ are present in healthy neurons, and regions of high neuronal loss have been found to lack them, suggesting that these inclusions may not, in fact, be the direct cause of toxicity [43].

A study of the formation of cytoplasmic inclusion bodies in Huntington's disease demonstrates that an increase in the amount of diffuse intracellular huntingtin correlates with cell death, while formation of inclusion bodies is actually protective, decreasing both toxicity and huntingtin levels elsewhere in the cell [70]. These large inclusions of unfolded protein within cells have typically been referred to as aggregates, and were believed to be the only sites of aggregates within cells. Recent work has found these inclusions are not the only type of aggregates that can be found in cells, however, as fibrillar aggregates have been observed within the cytoplasm outside of inclusion bodies [71] These aggregates are relatively small, ~100 nm in diameter and 1-2  $\mu\text{m}$  long, and are very similar to fibrils prepared by *in vitro* experiments. Cytoplasmic aggregates, contained both within and outside of inclusion bodies, present an alternative to the hypothesis that toxicity is caused by nuclear inclusions. It is also possible that

cytoplasmic oligomers (observed in [71]) contribute to toxicity in a manner similar to that proposed in other amyloidoses, in which small oligomers are believed to be the toxic species [29, 30].

Yet another proposed mechanism for polyQ toxicity, the toxic monomer hypothesis, suggests the polyQ monomer undergoes a toxic conformational transition that is more favorable at increased repeat lengths. This hypothesis is discussed in more detail in Section 1.3.2.2.

The work described here is based on the premise that aggregated structures of polyQ (rather than monomer) are important to disease manifestation. By characterizing the fundamental mechanisms of aggregation *in vitro*, we can gain a better understanding of the pathogenesis of polyQ disease.

## **1.3 AGGREGATION OF POLYGLUTAMINE CONTAINING PEPTIDES**

### **1.3.1 *In Vitro* PolyQ Aggregation**

Simple polyQ peptides have characteristic aggregation kinetics. When the aggregation pathway is presented as an aggregate concentration vs. reaction time graph, it typically assumes a sigmoidal shape [72]. This curve is representative of three dominating phases within the aggregation pathway. During the initial lag phase, no detectable aggregates are forming as monomeric polyQ begins to nucleate and elongate amyloid fibrils. Unlike a number of other amyloid forming peptides, polyQ appears to be resistant to secondary nucleation, so this phase is dominated by the formation of a primary nucleus and initiation of elongation [73]. Because secondary nucleation is not a contributing factor, the shape and duration of this lag phase can be

used to determine the size and stability of the amyloid nucleus, using a simple, straightforward mathematical model, as described below (1.3.5.1) [74]. Confirmation that this lag phase can be attributed to nucleation can be achieved through introduction of fibril seeds. By introducing preformed nuclei and maintaining a low monomer concentration, elongation of the introduced seed is energetically favored over spontaneous nucleation.

The second phase of polyQ aggregation is characterized by a steep increase in the sigmoidal curve. This phase is dominated by polyQ fibril elongation through monomer addition to the nucleated fibril. Aggregate concentration continues to increase until the third phase of the pathway, at which point monomer and aggregates slowly tend toward a dynamic equilibrium. The monomer component of reactions that have reached completion is not eliminated, but equilibrates as defined by the critical concentration,  $C_r$  (as described above, 1.1.2). When this equilibrium is reached, an aggregation reaction is described as having run to completion. These characteristic aggregation kinetics allow for the application of mathematical models that define this polymerization mechanism [22, 75]. One of these models is described in greater detail below (Section 1.3.5.1).

## **1.3.2 Structural Steps of PolyQ Fibril Formation**

### **1.3.2.1 The PolyQ Monomer Ensemble**

In an effort to learn more about intrinsically disordered proteins and the conformations they assume [76], numerous studies of monomeric polyQ peptides have been conducted. Monomeric polyQ peptides, both longer and shorter than the repeat length threshold in disease, are disordered in solution [64, 77, 78] and adopt compact coil structures in water [79-82]. This result is surprising because Gln residues are not hydrophobic, and would be expected to maintain

interactions with solvent in a random coil. FCS data indicates that a compact coil is favored; that is, aqueous solutions are poor solvents for polyQ peptides. The compact coil structure is assumed because chain-chain interactions are favored over side chain-solvent interactions. More specifically, side chain-backbone hydrogen bonds have been demonstrated to be the main stabilizing force in the polyQ compact coil [83]. Insolubility in water is a likely driving force behind the aggregation of polyQ fibrils, which form amyloid as a means of maximizing both peptide-peptide and peptide-solvent interactions.

According to the thermodynamic model of nucleation described in 1.3.5.1, any nucleus structure that forms from these compact coil monomers must form through an energetically unfavorable reaction. This reaction must either consist of protein folding steps which occur as the compact coil monomer assumes the structure of a monomeric nucleus, or some combination of folding and association that transpire as a multimeric nucleus is formed. When considered in terms of the energetics of folding [84], the compact coil monomer is stabilized by both the conformational entropy of its disordered state, as well as enthalpic contributions from hydrogen bonds and other stabilizing interactions that might occur within this conformation. In order to form the nucleus, it is necessary to first disrupt this compact coil structure—the existing side chain-backbone and water-backbone hydrogen bonds must be broken before the  $\beta$ -sheet stabilizing backbone-backbone hydrogen bonds essential to amyloid structure can be formed. Entropic contributions resulting from introducing order to a previously disordered structure must also be overcome. The nucleus likely has fewer stabilizing interactions than the compact coil, since it is unclear whether the side chains can be fully hydrogen bonded intramolecularly within the nucleus. Further, in the case of a multimeric nucleus, association of multiple molecules must occur, which require overcoming of a diffusional entropy barrier.

Through analysis of aggregation kinetics (monitored by sedimentation assay, 2.2.1) and application of a thermodynamic model for nucleated polymerization (1.3.5.1) to a series of peptides with repeat lengths ranging from Q<sub>23</sub>-Q<sub>27</sub>, it was confirmed that polyQ peptides of repeat length 26 or greater aggregate *via* a monomeric nucleus. Below this threshold, however, peptides aggregate *via* multimeric nuclei, with Q<sub>23</sub> assuming a tetrameric nucleus [18, 74, 85]. Understanding the underpinnings of this mysterious shift in nucleus size in the Q<sub>23</sub> to Q<sub>26</sub> region is one of the main goals of the research described in this thesis.

A  $\beta$ -hairpin has been proposed as a likely choice for the structure of the monomeric nucleus for polyQ aggregation [18, 19]. The  $\beta$ -hairpin structure is also consistent with structural data (described below, 1.3.2.4) that indicates it could be the building block of polyQ amyloid. This hypothesis has been justified by experiments in which  $\beta$ -turn encouraging mutations are introduced into short repeat length peptides. Insertion of D-Pro-Gly and Trp-zip motifs to the center and termini (respectively) of Q<sub>23</sub> analogs result in a decrease in nucleus size ( $n^*$ ) from 4 to 1 (1.3.2.5) [19]. These small modifications to the polyQ sequence lower the energy barrier that must be overcome by short repeat length peptides in order to assume a  $\beta$ -sheet rich, monomeric nucleus structure similar to the one formed by unbroken polyQ peptides with physiological repeat lengths. These data suggest that simple polyQ peptides are nucleated by and form fibrillar assemblies consisting of  $\beta$ -hairpins.

Even though these mutations enhance polyQ peptides' ability to nucleate through a  $\beta$ -hairpin structure and are expected to make the same energetic contributions to polyQ as they do in other peptides, they have no directly measurable effect on the structure of monomeric polyQ in solution, which exists in compact coil [19]. The lack of measurable impact on structure is due to the fact that the hairpin structure is so disfavored in the equilibrium between monomer and

nucleus that even a multi-fold improvement of nucleus stability results in an structural enhancement that cannot be directly observed but is only inferred in the nucleation kinetics.

### **1.3.2.2 The Toxic Monomer Hypothesis**

The toxic monomer hypothesis originated from studies in which polyQ specific antibodies bound to longer repeat length polyQ expansions more efficiently than shorter Gln repeats. The proposed explanation was that the higher repeat length peptides assumed a unique conformation that the shorter peptides did not [86]. 1C2, a polyQ conformational antibody, was found to preferentially bind to longer repeat lengths of polyQ, leading to the suggestions that it is the expansion of the Gln stretch in these polyQ peptides that results in adoption of a toxic conformation [87]. Later, the 3B5H10 antibody in particular appeared to preferentially bind an anti-parallel  $\beta$ -sheet conformation of polyQ that seemed to be formed by longer repeat length polyQ peptides [88]. Because long polyQ repeat expansions tend to more efficiently form  $\beta$ -sheet rich structure (i.e. fibrils), and because 3B5H10 appears to be a better predictor of neurodegeneration (compared to other polyQ binding antibodies) it was argued that this purported  $\beta$ -rich conformer is a toxic monomer species of polyQ [52, 87-89]. This claim was further rationalized by data interpreted to indicate that  $\beta$ -hairpin structures introduced to cells are toxic [90].

Disputing this hypothesis, the improved binding of some of similar antibodies to long polyQ peptides was proposed to be the result of a “linear lattice” model, in which longer polyQ binds to antibodies more efficiently because it has more ligand-binding sites than shorter, non-pathological polyQ sequences [91]. Additionally, new structural studies indicating that both long and short polyQ have similar structures, and that the 3B5H10 epitope is actually a short, linear, extended conformation of polyQ further discredit claims of the existence of a compact, toxic monomer [78, 92].

$\beta$ -sheet rich structures are a common feature of all polyQ fibrils (which are, by definition,  $\beta$ -sheet rich), and are simply formed more rapidly by longer repeat lengths, which have an increased tendency toward aggregation [74, 93]. PolyQ monomers of a variety of lengths, on the other hand, have been demonstrated to favor compacted coil, not  $\beta$ -sheet structures in aqueous solution (described in detail above, 1.3.2.1) [79]. Finally, no direct evidence of a link between any such conformer and toxicity has been presented that does not unequivocally rule out the formation of aggregates that then act as toxic species [88, 90].

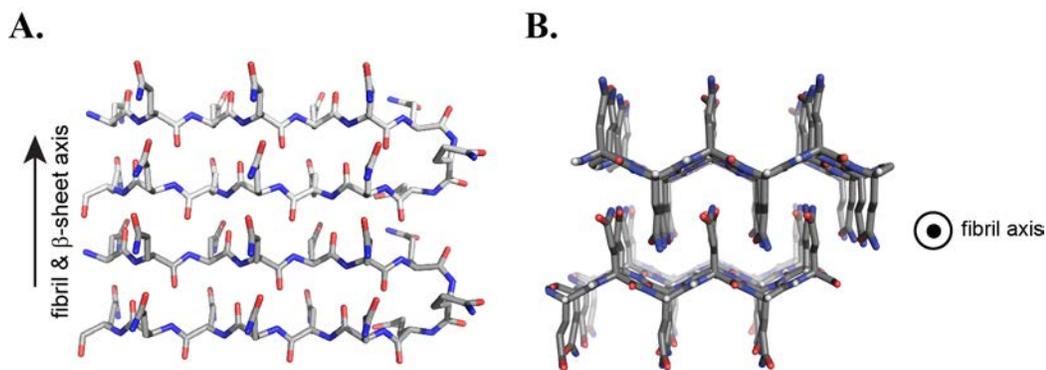
### **1.3.2.3 PolyQ Oligomers and Non-fibrillar aggregates**

Simulations of simple polyQ aggregation have suggested the formation of oligomeric species, within which the nucleus of amyloid formation has been proposed to form [94]. Similarly, *in vitro* experiments have resulted in the appearance of transient oligomers [95], but these structures have not been demonstrated *via* kinetic analyses to contribute to amyloid formation. It has been suggested that the appearance of these oligomers is due to differences in sample preparation [18], as polyQ oligomeric species have been observed under certain aggregation conditions [64, 65].

Along with changes in aggregation conditions, alterations to the simple polyQ peptide sequence have been found to result in oligomer formation. Addition of histidine residues to a simple polyQ peptide results in fibril formation *via* an oligomeric intermediate in acidic conditions but not at neutral pH [73]. While the presence of these oligomers is associated with slowed aggregation kinetics, they have not been demonstrated to contribute to the nucleation mechanism. Htt-exon 1 model peptides are capable of aggregating *via* an oligomeric intermediate mediated by Htt<sup>NT</sup> (described in detail 1.3.3.2) [96]. There is no experimental evidence at this time for polyQ mediated interactions leading to kinetically important oligomer formation.

### 1.3.2.4 PolyQ Amyloid

PolyQ peptides generally form  $\beta$ -sheet rich, amyloid-like structures when aggregated. X-ray fiber diffraction studies of fibrils formed by polyQ peptides with repeat lengths ranging from 9-45 demonstrate that, regardless of repeat length, all polyQ fibrils are comprised of slab-like  $\beta$ -sheet structures [97]. One interpretation of x-ray diffraction data proposes that  $\beta$ -strands of polyQ peptide aggregate into anti-parallel  $\beta$ -sheet with interchain hydrogen bonds (between both backbone and side chain amides) in the direction of fibril growth [20]. Gln side chains interdigitate across the  $\beta$ -sheet in a direction perpendicular to the fibril axis (**Figure 1-4 B**). This structural model supports the idea of  $\beta$ -hairpins as building blocks of polyQ fibrils (**Figure 1-4 A**). In addition, CD, FTIR and solid-state NMR (ssNMR) data of both simple polyQ and htt-exon 1 model peptide fibrils confirm  $\beta$ -sheet rich structures with rigid, solvent protected cores [15, 96, 98].



**Figure 1-4. Structure of PolyQ  $\beta$ -strands.**

A  $\beta$ -hairpin model for polyQ (A), this structure is believed to act as both nucleus and building block within amyloid fibrils. Gln side chains interdigitate to enhance stable elongation as they form  $\beta$ -sheet (B) From CH and PVW, data from [20].

To test whether a multistrand  $\beta$ -sheet was a good structural model for polyQ amyloid formation, multiple L-Pro-Gly or D-Pro-Gly motifs were introduced to long polyQ peptides [99].

L-Pro-Gly insertions were introduced at Gln segment lengths of 7, 8, 9, or 10, to create the following sequence  $K_2Q_N(PGQ_N)_3K_2$  (referred to as PGQ peptides). The  $Q_9$  and  $Q_{10}$  containing peptides aggregated quickly, at rates comparable to an uninterrupted  $Q_{45}$  peptide, supportive of the hypothesis that polyQ peptides form a four stranded, antiparallel  $\beta$ -sheet. The shorter peptides aggregated considerably more slowly, most likely because the uninterrupted Gln chain length is not capable of efficiently forming  $\beta$ -sheet. Since the L-Pro-Gly based turns are believed to require incorporation of a Gln on either side (to result in a four residue turn, -QPGQ-), the shortest amino acid stretch that supports efficient  $\beta$ -sheet formation must contain seven Glns exclusive of those involved in turns (such as in the  $PGQ_9$  peptide). Substitution of D-Pro-Gly for L-Pro-Gly in the  $Q_9$  context further accelerates aggregation, most likely due to an improved ability to form two residue turns, in particular  $\beta$ -hairpins [100]. When the extended  $\beta$ -sheet in the PGQ peptides is broken by Pro insertions, aggregation kinetics are significantly diminished and sometimes eliminated, confirming that formation of stable extended  $\beta$ -sheet is essential to fibril formation.

This four-strand model was further tested with an htt-exon 1 model protein expressed in mammalian cells [101]. As in the previously described experiments, PGQ constructs were used to induce an alternating  $\beta$ -strand, turn structure in the polyQ region of htt-exon 1 proteins. The htt-exon 1 containing  $PGQ_9$  repeats aggregated significantly and was toxic to cells, while the same protein with Pro residues interrupting the Gln stretches was stably expressed but resulted neither in detectable aggregates nor toxicity. These results suggest that the ability to form a  $\beta$ -sheet rich, amyloid-like aggregated structure is critical to induce toxicity of polyQ sequences.

### 1.3.2.5 $\beta$ -turn inducing mutations

Experimental methods to test the nature and roles of  $\beta$ -hairpin structures in polyQ aggregates necessitate the use of  $\beta$ -hairpin encouraging mutations. Some of these, such as L-Pro-Gly and D-Pro-Gly motifs, have been discussed, but there are a number of other  $\beta$ -hairpin encouraging mutations, some of which have been incorporated into the sequences listed in **Table 1-3** [100, 102-106].

**Table 1-3. Hairpin encouraging mutations affect n\*.**

Peptides	K2Q23K2	K2Q10PGQ11K2	D2Q23K2	K2Q10pGQ11K2	K2CQ22CK2h	AcWQ11pG Q11WTGK2
n*	3.9	1.4	1.6	0.7	0.8	0.9

Peptides	AcWQ16WTGK2	AcWQ18WTGK2	AcWQ21WTGK2	AcWQ22WTGK2
n*	4.6	4.4	3.1	2.7

Modified from [19].

Experiments using disulfide bonds, Trp-zip motifs and attractively charged residues at the termini of Q<sub>23</sub> analog proteins all enhanced aggregation and decreased nucleus size from a tetramer (**Figure 1-5, Table 1-3**) [19]. Disulfide bonds, addition of D-Pro-Gly residues to the center of the sequence, and the combination of a Trp-zip with D-Pro-Gly all resulted in aggregation through a monomeric nucleus. The D-Pro-Gly sequence, however, is the best choice for any potential studies of the roles other flanking sequences may have on aggregation of monomeric polyQ, as it is the only mutation to stabilize a monomeric nucleus with no change in the terminal sequences of the peptide.

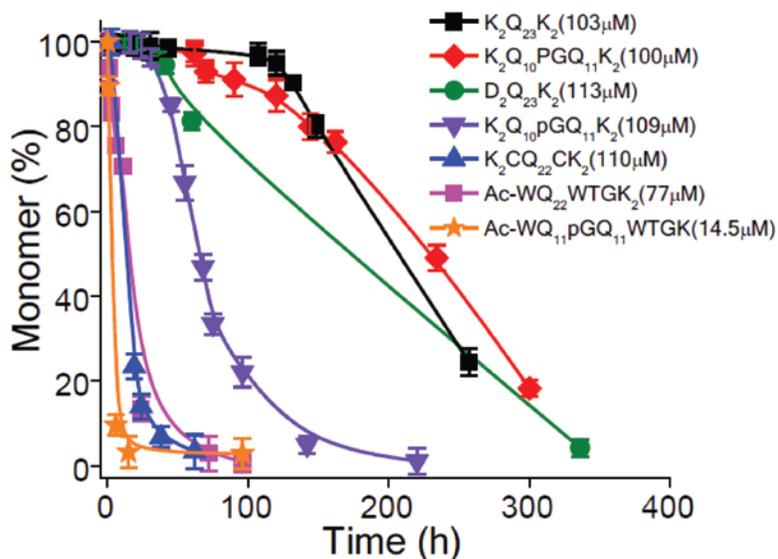


Figure 1-5. Aggregation kinetics of hairpin encouraging peptides [19].

### 1.3.2.6 Summary

Taken as a whole, these structural data of polyQ peptides build a basic picture of the transitions a molecule must undergo as it evolves from disordered monomer to a highly ordered and rigid amyloid fibril. As we gain more information regarding the thermodynamics and kinetics of fibril assembly and initiation, this structural information will be important to the development of a clearly defined mechanism of aggregation. By combining the existing knowledge of polyQ aggregation properties and structure with new data describing the biophysical contributions of a peptide's primary structure to its aggregation, this work clarifies the impact various structural features have on polyQ aggregation.

### 1.3.3 Flanking Sequences in PolyQ Aggregation

#### 1.3.3.1 Solubilizing charges are required to study simple polyQ aggregation

*In vitro* studies of polyQ disease proteins make use of either chemically synthesized or recombinantly expressed peptide models. Due to their insolubility in water, simple polyQ peptides require addition of solubilizing charges to their termini [77, 107, 108]. Addition of positively charged Lys residues as a solubility aid is common practice when working with poorly soluble, hydrophobic peptides [109]. Thus, nearly all *in vitro* studies of simple polyQ peptides have been conducted with a pair of Lys residues on both N- and C-termini.

When pure polyQ sequences proved insoluble in the first *in vitro* studies of polyQ peptides, a 15-residue Gln sequence was solubilized with four charged residues, two Asp on the N-terminal end, and two Lys on the C-terminus [108]. Once the polyQ peptide was successfully solubilized, it was possible to generate aggregates and study their structure. This work found that solubilized polyQ peptides aggregate into highly stable,  $\beta$ -sheet rich structures, likely consisting of  $\beta$ -hairpins with strong hydrogen bonding.

Subsequent studies of a similar peptide ( $K_2Q_{15}K_2$ ) revealed that while soluble in aqueous solution, it still formed aggregates that could be pelleted out by ultracentrifugation [77]. Addition of longer Ala and Lys containing sequences to the N- and C-termini led to complete solubilization of both long and short (17 and 9 Glns, respectively) polyQ peptides. When completely solubilized, these polyQ with short expansions were found by CD to be soluble random coil.

Other studies indicate that, at low concentrations, simple  $Q_{22}$  (with no flanking residues) is soluble, and that it forms  $\beta$ -sheet rich structures [110], rather than random coil as previously proposed [77]. The conclusion was that the ability of a given polyQ peptide sequence to form

large,  $\beta$ -sheet rich structures is what determines its propensity to aggregate, and that interruptions to the polyQ sequence impede the ability of polyQ to form  $\beta$ -sheet might inhibit disease progression.

The disagreement between these two studies ([110] and [77]) regarding whether simple polyQ monomer exists as random coil or  $\beta$ -sheet in aqueous solution raised a number of fundamental questions regarding the approach to *in vitro* studies of polyQ. In [111], Altschuler argued the importance of ensuring the starting material is monomeric, and claims there is insufficient data to justify the assertion that polyQ monomer forms  $\beta$ -sheet [110]. Sharma counters that Altschuler's use of large, non-physiological flanking sequences to solubilize polyQ (in [77]) may have played a role in the observation of random coil as the structure of short polyQ peptides, and suggests a series of more physiologically relevant flanking sequences [112]. Both arguments raise valid points: the first is that to make accurate arguments about the polyQ monomer structure and the aggregation pathway, it is critical to first guarantee a fully disaggregated monomer as the starting material. The second is that it is important to consider how the sequence context of polyQ might influence the peptide's properties before using their behavior to draw more generalized conclusions about simple polyQ expansions.

In an effort to ensure that all starting material for polyQ peptide studies is monomeric and entirely disaggregated, our lab developed a disaggregation protocol [107], described in detail in section 2.1.2. This disaggregation protocol was combined with tests of a series of flanking amino acids to determine the optimal choice for solubilizing relatively long polyQ peptides (~30 Glus) [65]. The use of flanking charges has a notable and dramatic effect on the solubility of polyQ peptides, but no discernible effect on the final morphology of the resulting aggregates. Lys pairs were chosen as the optimal solubilizing flanking residues, as they appeared to have the greatest

solubilizing effect, resulting in decreased aggregation propensity of Q<sub>15</sub> when compared to D<sub>2</sub>Q<sub>15</sub>K<sub>2</sub> and DKQ<sub>15</sub>DK. Increasing solubility of the polyQ sequence with these charged residues allowed for more detailed studies of longer, more physiologically relevant repeat lengths than had been previously conducted.

### **1.3.3.2 Disease relevant flanking sequences can impact aggregation mechanism**

Peptide models of huntingtin (htt) exon 1 have demonstrated that both the N-terminal 17 amino acids (Htt<sup>NT</sup>) and the C-terminal poly-Pro region play important roles in modulating aggregation. When their impacts on polyQ aggregation are considered individually these sequences have interesting effects. Addition of a poly-Pro segment 6-10 repeats long to the C-terminus of polyQ causes a marked decrease in the rate of aggregation and stability of the resulting fibrils [51]. Use of only 3 Pro residues or attachment of a poly-Pro (of length 10) to the N-terminus of the peptide did not have any observable effects on polyQ aggregation. These C-terminal poly-Pro modulations had no effect on the calculated nucleus size and were found to hinder aggregation of A $\beta$  in a similar manner. It is believed that the decreases in aggregation rate and fibril stability are due to a change in the conformational ensemble so it contains fewer aggregation competent monomer conformations with the Pro extension than without it. These effects are of particular interest because of the presence of both N- and C-terminal Pro-rich regions in a number of polyQ proteins (as mentioned in 1.2.1).

Incorporation of the Htt<sup>NT</sup> sequence to either the N- or C-terminus of polyQ results in a significant acceleration in aggregation rate [96]. When both Htt<sup>NT</sup> and a poly-Pro region are added to polyQ (to the N- and C-termini respectively), the enhancement effects of the Htt<sup>NT</sup> sequence dominate—this htt exon 1 model peptide aggregates faster than the same length simple polyQ. In spite of the apparent dominance of the Htt<sup>NT</sup> sequence, Htt<sup>NT</sup>-Q<sub>N</sub> peptides still

demonstrate a repeat length dependence similar to that of simple polyQ. Work to determine whether this repeat length dependent aggregation of exon 1 model peptides is related to modulating effects of Htt<sup>NT</sup>, the repeat length dependence of simple polyQ (discussed below in 1.3.4), or both is ongoing.

The full htt exon 1 model peptide, containing both Htt<sup>NT</sup> and poly-Pro regions, is capable of aggregating through a two-step pathway [96]. In the first step, modulated by Htt<sup>NT</sup>, the N-termini associate to form oligomers. Then, since the local polyQ concentration is enriched, the expanded polyQ chains exhibit a greatly enhanced ability undergo conformational changes to nucleate amyloid, and  $\beta$ -sheet rich, amyloid fibril formation is initiated. The end products of this reaction are amyloid fibrils in equilibrium with exon 1 monomers: any remaining non-amyloid oligomers resulting from the first step of this mechanism are expected to dissociate back into monomer. Thus, addition of a flanking sequence as short as 17 amino acids can result in a dramatic change to the aggregation mechanism.

Ataxin 3 is also capable of forming fibrils *via* a mechanism with two distinct steps [113]. The first, governed by the N-terminal Josephin domain, occurs regardless of Gln repeat length, and results in the formation of SDS sensitive amyloid. The second part of the pathway is dominated by polyQ interactions and results in formation of a more stable, SDS resistant fibril. In both of these systems, polyQ mediated aggregation acts as the key step in a series leading to the formation of amyloid such as that observed *in vivo* (while the non-polyQ mediated steps are important, they do not result in the formation of aggregates comparable to those found in disease states).

The presence of a polyQ mediated aggregation mechanism even within the context of disease relevant peptides is significant, and indicates that furthering our understanding of the simple polyQ aggregation pathway is essential to an understanding of these disorders.

### 1.3.4 PolyQ Demonstrates Repeat Length Dependent Aggregation

Repeat length dependent aggregation of polyQ *in vitro* correlates well with the repeat length dependence of pathogenicity [74]. Studies of recombinant exon 1 confirmed the repeat length dependent aggregation of polyQ both *in vitro* and *in vivo* for repeat lengths ranging from 20 to 51 [63, 114] (with repeat lengths less than 32 forming no detectable aggregates) and presented data that suggested *in vitro* aggregation is consistent with a nucleation dependent mechanism. For repeat lengths greater than 32, aggregation of exon 1 was dependent on both protein concentration <sup>2</sup> and time—a specific concentration of monomer was required to initiate aggregation, and aggregation was always preceded by a lag-phase, the length of which decreased with increasing repeat length. Furthermore, addition of a preformed seed eliminated the lag phase, all features consistent with a nucleated growth mechanism.

Synthetic simple polyQ peptides also demonstrate repeat length dependent aggregation [64, 74]. PolyQ peptides with the sequence  $K_2Q_NK_2$ , where  $N = 5, 15, 28, 36, 42, 44,$  and  $47$  were synthesized, disaggregated, and their aggregation kinetics were monitored *via* ThT binding or sedimentation assay (methods described in 2.2.1 and 2.2.2). The  $Q_5$  and  $Q_{15}$  peptides underwent little detectable aggregation during the course of these experiments (though  $Q_{15}$

---

<sup>2</sup> Amount of aggregate was determined by filtration to isolate aggregates followed by binding with an anti-AG51 antibody (specific to exon 1 aggregates) *via* dot blot and subsequent quantification.

aggregation has been detected with extremely long incubation periods). The larger peptides demonstrated features of nucleated growth, with lag phases that decreased with increases in Gln count, and monomeric nuclei. Fibril stability as measured by  $C_r$  (critical concentration) also varied with repeat length: short peptides had high  $C_r$  values ( $C_r$  for Q<sub>15</sub> was ~30  $\mu$ M), while the  $C_r$  was notably decreased for the longest repeat lengths ( $C_r$  for Q<sub>44</sub> was ~70 nM). These results suggest that increasing repeat length leads to increasingly favorable polymerization.

PolyQ peptides of repeat length 26 or greater have been demonstrated to aggregate *via* a monomeric nucleus, while shorter peptides have been found to aggregate through multimeric nuclei, indicating a clear relationship between repeat length and nucleus size of simple polyQ peptides (1.3.2.1) [18]. This switch in nucleus structure might explain the repeat length dependence of polyQ in the 20's, but leaves the basis for repeat length dependence in longer peptides unclear.

Repeat length also appears to play a role in fibril structure: vibrational spectroscopy studies of polyQ peptides ranging in repeat length concluded that while all polyQ fibrils have similar secondary structures (as observed by vibrational circular diocroism), there are detectable morphological differences in the structures formed by polyQ with repeat lengths greater than 35 as compared to those with 30 or fewer Gln residues [115]. Hydrogen exchange data indicates that longer repeat length polyQ fibrils exhibit a less protected core than shorter fibrils, in spite of increased fibril stability, suggesting differences in the core  $\beta$ -sheet structure that makes up these fibrils. The proposed threshold for these structural differences again lies between Q<sub>30</sub> and Q<sub>35</sub>, similar to the repeat length threshold commonly observed in disease.

The *in vitro* effects of repeat length on both aggregation and structure may have implications on the pathological repeat length threshold observed in disease. Could it be that

polyQ diseases only manifest when repeat length surpasses 30 because of the challenges involved in initiating nucleation when a monomeric nucleus is not favored? Aggregate structure may also play a pathogenic role, as it also appears to have a repeat length dependence as detected through fibril stability and spectroscopic measurements. This leaves us questioning whether repeat length dependence of pathogenicity is a result of the influences of Gln count on the aggregate structure, or on the rate and degree of aggregation of polyQ peptides.

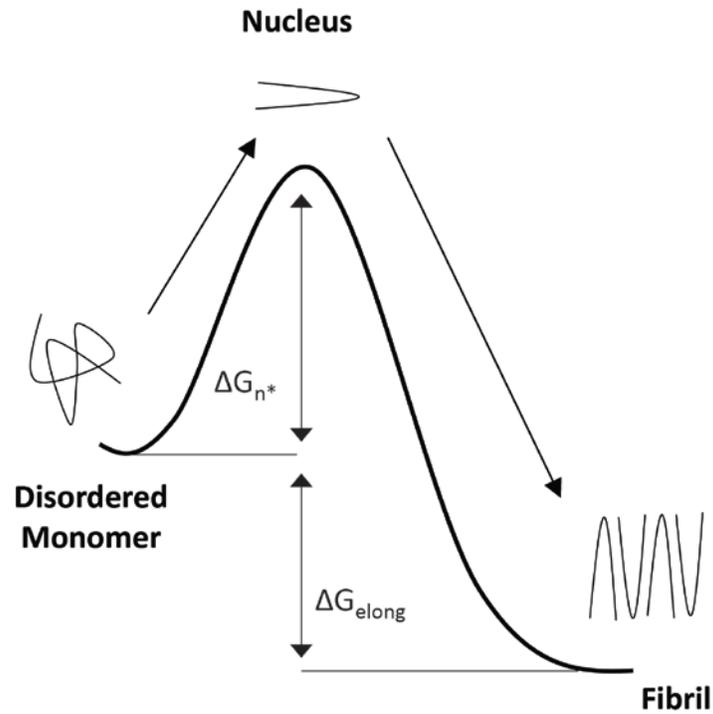
### **1.3.5 Mathematical Models for PolyQ Aggregation**

#### **1.3.5.1 A Nucleated Polymerization Model of Aggregation**

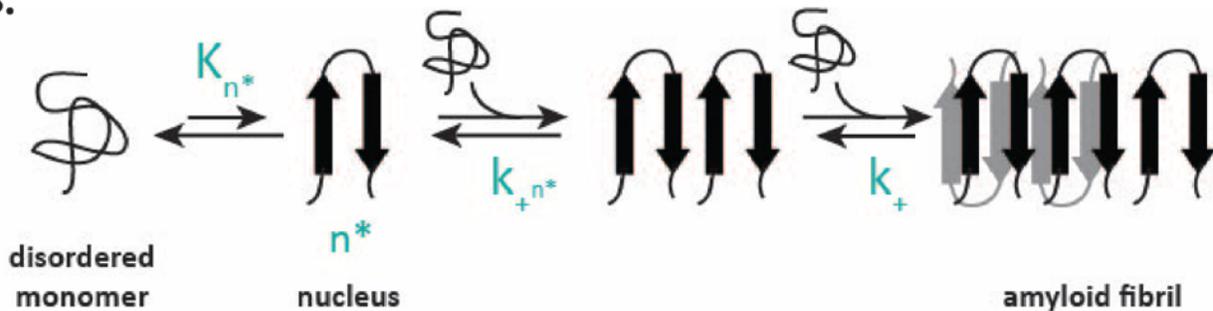
In spite of extensive study, there is still much that remains to be learned about polyQ aggregation and amyloid formation. In particular, the initiation of fibril formation is still largely a mystery. Developing an understanding of this critical nucleation step is essential to our understanding of amyloid fibrils.

According to a thermodynamic nucleated polymerization model for aggregation developed by Eaton and Ferrone [22], the thermodynamic nucleus exists as a maximum energy transition state that the system must pass through in order for spontaneous aggregation to occur (**Figure 1-6**). Because, by definition, the nucleus is the highest energy structure on the aggregation pathway, it is also the least stable and least populated structure along the path to amyloid formation, further challenging any studies of its nature. Its crucial position as a rate limiting step in the aggregation pathway makes studies of the nucleation phase of amyloid formation essential to our understanding of amyloid fibrils themselves. Furthermore, a thorough knowledge of amyloid nucleation may allow for treatment of disease by preventing initiation of

A.



B.



**Figure 1-6. Thermodynamic model of polyQ aggregation.**

A simple free energy diagram of polyQ aggregation (A). The nucleus exists as an unstable, high energy state, resulting in its role as a rate limiting step in the folding pathway. Two free energies describe this pathway, the free energy of nucleation ( $\Delta G_{n^*}$ ) and the free energy of elongation ( $\Delta G_{\text{elong}}$ ). Only the molecules which form the amyloid nucleus must surpass the  $\Delta G_{n^*}$ , after the nucleus is formed, the energy barrier is lowered for elongating molecules. An illustration of the steps in the polyQ fibrillization mechanism (B). Disordered monomer folds into a nucleus, which elongates by monomer addition into a fibril. In order to generate his simple mechanistic model, it was assumed by Ferrone that elongation of the nucleus and fibril proceed identically, and that  $k_+ = k_{+n^*}$ .

fibril formation. Development of an experimental model of the nucleation mechanism may make it possible to finally design and test therapeutics that target amyloid fibril nuclei.

The thermodynamic model of homogeneous nucleation developed by Eaton and Ferrone has been successfully applied to polyQ aggregation [22, 74, 116, 117]. According to our application of this model, the polyQ amyloid nucleus exists in an equilibrium state with the compact coil monomer, with the compact coil structure heavily favored (**Figure 1-6B**). For simple polyQ peptides, elongation of the nucleus occurs through monomer addition, a much more energetically favorable process than nucleus formation.

Early phase aggregation kinetics are analyzed to determine nucleus size ( $n^*$ ), as well as a complex parameter ( $\log(\frac{1}{2}k_+^2K_{n^*})$ ) that relates the nucleation equilibrium constant ( $K_{n^*}$ ) and rate constant for fibril elongation ( $k_+$ ). This complex parameter is indicative of a peptide's ability to undergo spontaneous aggregation, and can be used to compare the impacts changes to the polyQ sequence have on aggregation efficiency.

Further analysis helps us untangle the  $\log(\frac{1}{2}k_+^2K_{n^*})$  parameter. The pseudo-first order rate constant ( $k^*$ ) is measured by fitting seeded elongation reaction kinetics. When an appropriate amount of seed is used, a constant number of elongation sites are maintained, so elongation is solely dependent on monomer concentration. Biotinyl-polyQ titration experiments make it possible to determine of the amount of growing ends in a seed sample. Concentration of growing ends and  $k^*$  are used to determine the second order elongation rate constant,  $k_+$  which corresponds to the rate at which monomer adds to the growing fibril. Determination of  $k_+$  allows for isolation of  $K_{n^*}$  from the complex parameter mentioned above [85]. The free energy of nucleation ( $\Delta G_{n^*}$ ) can be calculated from this nucleation equilibrium constant. These approaches are discussed further in Methods (2.1.5, 2.2.4, 2.2.5, 2.2.6) and Chapter 3.0.

In Ferrone's mathematical treatment of the Eaton and Ferrone model, in order to simplify equations to make possible an approximate solution,  $k_{+n^*}$ , the rate at which monomer adds to the nucleus is assumed to be identical to  $k_+$ , the rate at which monomer adds to fibrils. This equality is justified by the argument that on-rates are predominantly defined by long-range interactions, such as diffusion, with electrostatics possibly having some effect. These long-range forces are not expected to change much as the nucleus grows into a fibril, and so it is possible to justify  $k_{+n^*} \equiv k_+$ .

In addition, our lab has found that by monitoring reactions as they run to completion and conducting dissociation reactions it is possible to determine the critical concentration ( $C_r$ ) and from it the free energy of elongation ( $\Delta G_{\text{elong}}$ ) for a given peptide (as described in 1.1.2).

These thermodynamic and kinetic values characterize both the nucleus and final amyloid product, and allow us to define key steps in fibril formation. The ability to extract thermodynamic quantities specific to each peptide sequence allows for the determination of the importance of various structural features in amyloid assembly [19]. The mathematical methods used in these calculations are explained in detail in the Methods chapter (2.2.6) and based on those described and derived in [22]. Our use of this model is further justified by its excellent fit to our polyQ data, as illustrated in later chapters of this work (see **Figure 4-7**, **Figure A-1**, **Figure A-9**, **Figure A-11**, **Figure A-15**). We also never see the products of heterogeneous aggregation mechanisms, either by EM or other methods such as DLS [18], confirming our choice of a model of homogeneous nucleated aggregation mechanism to fit simple polyQ aggregation.

### 1.3.5.2 Alternate Models of Aggregation

The above model is not the only mathematical treatment of aggregation and nucleation. The Weissman group developed a mathematical model which fits the aggregation kinetics of Sup35, a yeast prion protein [118]. Secondary nucleation plays a significant role in the aggregation mechanism of this protein. By modifying the Ferrone model described above, it was possible to generate a model consistent with amyloid nucleation through a nucleus of  $n^* \leq 3$ . Elongation then occurred by subsequent monomer addition, followed by secondary nucleation through fragmentation [119]. Though oligomers were observed during the aggregation of Sup35 [34], elongation appears to be dominated by monomer addition, not association and rearrangement of oligomeric species, at least at the relatively low concentration of the Weissman experiments. Fortunately, simple polyQ aggregation does not undergo secondary nucleation, so an alternative approach such as this is not necessary for our studies.

An alternative to the model of Chen et al. (described in 1.3.5.1) [74] has been proposed based on a heterogeneous model of polyQ aggregation [120]. This model assumes the presence of on-pathway oligomeric intermediates that undergo conformational conversion into a nucleus structure, which is then elongated by monomer or oligomer addition. As discussed previously (1.3.2.3), however, there is no experimental evidence of on-pathway oligomers in the simple polyQ aggregation mechanism, and application of the previously described homogeneous model of polyQ aggregation mechanism has been confirmed to be an equally good fit in estimating values such as nucleus stability [121].

## **1.3.6 Fundamental Questions Being Addressed**

### **1.3.6.1 Flanking sequences in polyQ peptide aggregation**

The use of flanking charges in studies of polyQ peptides has become commonplace (1.3.3.1). Simple polyQ peptides are insoluble and spontaneously aggregate upon introduction to aqueous solution at pH 7, but addition of 4 Lys residues (2 at each terminus) solubilizes the monomer and allows for the study of aggregation kinetics [65]. Concerns have been raised regarding the influence of introducing a strong positive charge to the termini of polyQ peptides [122], and the role the electrostatics of these amino acids play in the polyQ aggregation mechanism is unclear. Small additions to the polyQ peptide sequence can have major impacts on peptide aggregation properties [19, 99]. The difference in physiological repeat length between different CAG-repeat diseases, and recent studies of htt-exon 1 model peptides suggest that the sequences flanking the polyQ region are important in modulating the mechanism of polyQ aggregation [51, 96, 123, 124]. Consideration of these facts leads to a series of important questions: how many charges are necessary to solubilize polyQ? Is fibril morphology affected by the magnitude of the net charge? Under what circumstances do flanking sequences have a significant impact on aggregation mechanism or ability? How might the impact of flanking sequences vary with a change in repeat length? Do flanking sequences affect one portion of the aggregation mechanism more significantly than others?

### **1.3.6.2 Repeat length dependence**

The repeat dependence of polyQ disease manifestation has been previously discussed (1.2.2). Similar repeat length roles have been identified in *in vitro* studies of polyQ, in which the size of the polyQ expansion plays a role in both elongation kinetics and nucleation of amyloid [18, 74].

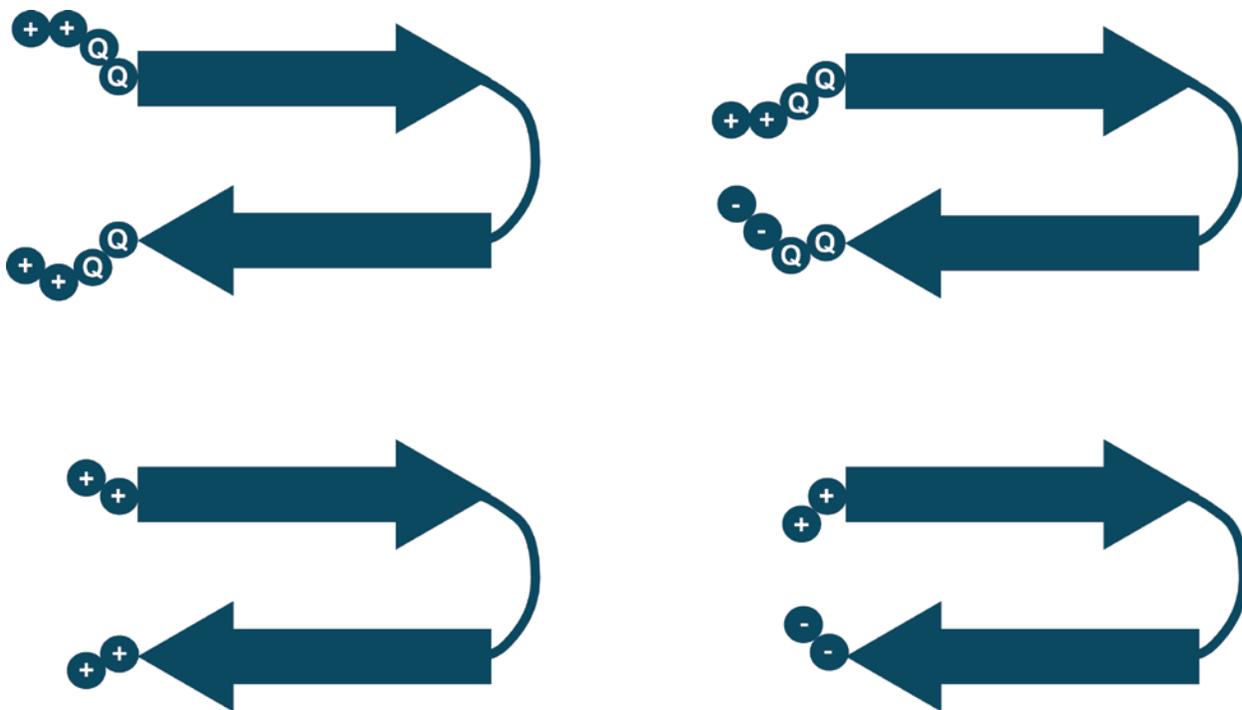
In spite of the clear importance of repeat length both in the disease state and experimental studies, little has been done to discern the biophysical implications of changes in Gln content. Will a more detailed study of repeat length thermodynamics confirm previous reports of correlations between fibril stability and repeat length? What trend exists between stability and repeat length (i.e. will there be discontinuities)? What role does repeat length play in nucleation efficiency? Is there an energetic explanation for why short repeat lengths (such as Q<sub>23</sub>) are not observed to aggregate *via* a monomeric nucleus? Is there an energetic explanation for why aggregation rates continue to increase as repeat lengths increase, even among the family of repeat lengths which aggregate *via* a monomeric nucleus?

## 1.4 PRELIMINARY STUDIES

### 1.4.1 Peptide Design

#### 1.4.1.1 Optimization of peptide length

To measure the quantitative and qualitative effects of charged flanking sequences, it is essential to first develop the appropriate polyQ core context. This work utilizes two polyQ frameworks for the studies of flanking charges. The guiding factor in the selection of these core sequences was the understanding that peptides exhibiting monomeric aggregation nuclei are simplest to interpret, since nucleation in this case is reduced to a protein folding problem, rather than a combination of folding and association [51].



**Figure 1-7. Positive and negative effects on hairpin closure due to terminal charge interactions.**

Use of repulsive terminal charges on termini can be expected to lead to ‘fraying’ of the termini (left), and act as a destabilizing force in the closure of a  $\beta$ -hairpin structure. If there are more residues than the absolute minimum necessary to stabilize a closed hairpin (top), this effect can result in minimization of the interactions between the termini. Use of attractive charges on termini (right) supports hairpin closure.

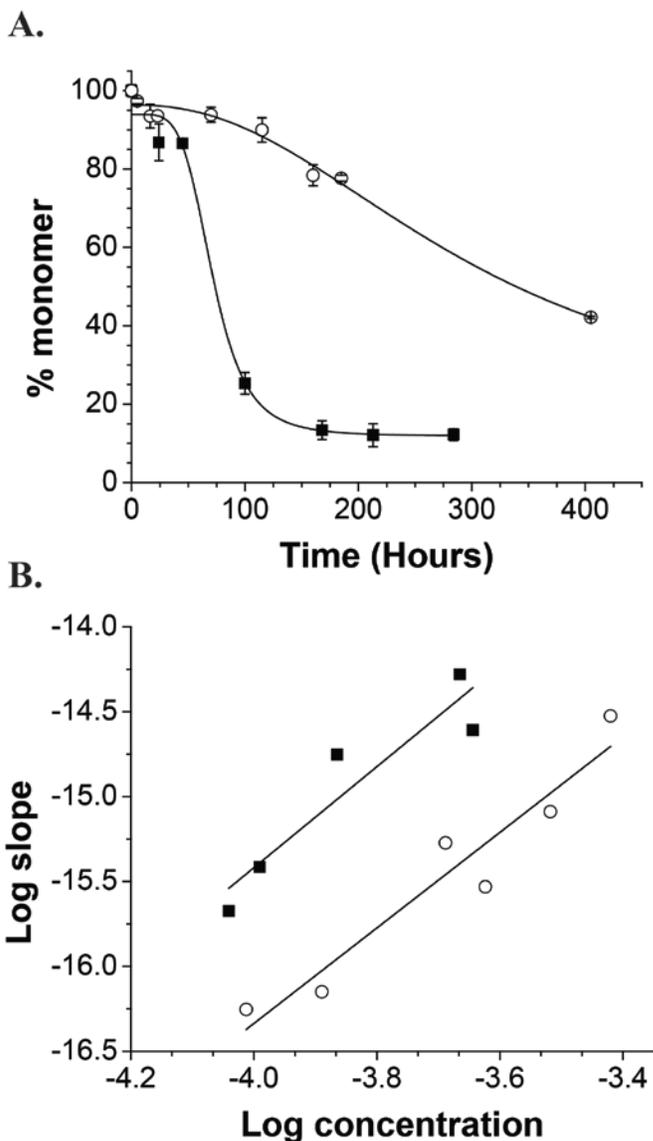
To measure quantitative changes and maximize charge interactions between the N- and C-termini during both fibril elongation and nucleation, a polyQ sequence containing a D-Pro-Gly  $\beta$ -hairpin favoring mutation was selected. This peptide was chosen to ensure monomeric nucleation in a short polyQ, while minimizing any mechanism changes that might be introduced by variations in flanking sequence charge. To ensure that the influence of terminal charges on aggregation was accurately measured, it was necessary to use a peptide that minimized separation of the termini and flexibility in the hairpin structure. Optimization of the number of residues that contributed to stabilizing interactions within the hairpin by decreasing the overall Gln repeat length was expected to accomplish this goal. When decreasing Gln repeat length, it is

also important to consider peptide aggregation time: as repeat length is decreased, aggregation time increases significantly, eventually (at low enough repeat lengths) making it difficult to obtain quality, interpretable data. To determine the shortest polyQ  $\beta$ -hairpin peptide with an  $n^*$  of 1 (or monomeric nucleus) that can be feasibly studied, two peptides were considered:  $K_2Q_{10}P^D GQ_{10}K_2$  and  $K_2Q_9P^D GQ_9K_2$  (**Figure 1-8**).

Both peptides were found to aggregate *via* monomeric nuclei, though aggregation of the  $Q_9$  based peptide was significantly slower than aggregation of the  $Q_{10}$  peptide. While  $Q_8$  or even  $Q_7$  based peptides might maintain a monomeric nucleus (based on data in [99], only seven Gln residues capable of interacting are necessary to form stable  $\beta$ -sheet), their aggregation kinetics would be prohibitively slow when initiated at feasible (based on amounts of peptide that can be acquired) concentrations.

Of particular note is the significant influence the D-Pro-Gly insertion has on otherwise short Gln repeat lengths. Normally, peptides  $Q_{25}$  and shorter have multimeric nuclei, but in the case of  $K_2Q_9P^D GQ_9K_2$ , a  $Q_{20}$  analog peptide is still monomeric because of the influence of the  $\beta$ -hairpin encouraging mutation. Because both the  $Q_{10}$  and  $Q_9$  peptides have monomeric nuclei, the slopes of the fit lines in the log-log plot are parallel (**Figure 1-8 B**). The significant displacement in the data between the two peptides corresponds to the notable differences in kinetics as illustrated in (**Figure 1-8 A**), and suggests that they are not due to a change in nucleus size, but instead an effect of peptide length on elongation efficiency and nucleus stability (as indicated by the y-intercept of the log-log plot). The magnitude of this efficiency difference, as measured in the difference in the y-intercepts of the log-log plots, plays an important role in validating the approach to understanding repeat length effects described in Chapter 5.0. Otherwise, this

experiment resulted in our choice of the Q<sub>9</sub> peptide as one of the model systems for the study of charge effects (Chapter 4.0).



**Figure 1-8. Aggregation of K<sub>2</sub>Q<sub>10</sub>P<sup>D</sup>GQ<sub>10</sub>K<sub>2</sub> and K<sub>2</sub>Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K<sub>2</sub>.**

The kinetics of K<sub>2</sub>Q<sub>10</sub>P<sup>D</sup>GQ<sub>10</sub>K<sub>2</sub> (■) and K<sub>2</sub>Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K<sub>2</sub> (○) at concentrations of 125.6 μM and 128.9 μM, respectively (A). In spite of the similarities in reaction concentration, the Q<sub>10</sub> peptide aggregation rate is significantly faster than that of the Q<sub>9</sub> peptide. The log-log plots of these two peptides reveals nearly identical slopes, and therefore similar n\* values (B). The Q<sub>10</sub> peptide has an n\* of 0.98, and the Q<sub>9</sub> peptide n\*=0.81. Both of these values are consistent with those typically observed for monomeric nuclei.

## 2.0 EXPERIMENTAL METHODS

### 2.1 PEPTIDE PREPARATION AND REACTION PROTOCOLS

#### 2.1.1 Peptide Synthesis and Purification

All peptides used in these studies were synthesized by solid-phase synthesis and obtained crude from the Keck Biotechnology Center (<http://medicine.yale.edu/keck/ssps/index.aspx>) in synthesis scales of 25-100  $\mu$ M. All crude peptides were stored at -20°C for the short term and -80°C for long term prior to purification.

Crude peptides were purified using reverse phase chromatography on an Agilent Zorbax C3 column connected to a Biorad Biologic Dual Flow system. Peptides were dissolved in 100% formic acid (Sigma) for 2-3 minutes, until completely dissolved. The peptide-formic acid mixture was then diluted to 20% formic acid with deionized water prior to injection onto the reverse phase column. It is important to be sure that the peptide is injected shortly after having been dissolved in formic acid to prevent formylation of the sample. Any formylation that does occur can be isolated *via* the purification step and would be detected by MS if any remained with the purified peptide sample. The buffers used for purification were water (Fisher Scientific, HPLC grade) with 0.05% trifluoroacetic acid (TFA) (Sigma) as buffer A, and acetonitrile (Fisher Scientific, HPLC grade), also with 0.05% TFA (Sigma) as buffer B. To ensure a clean column

prior to peptide injections, “blank runs” of 20% formic acid injections were performed prior to each peptide purification run. All purification methods maintained a solvent flow rate of 4 mL/minute with a buffer B gradient optimized to the peptide being purified. Eluted fractions were collected using a Biorad fraction collector in volumes of 0.3-0.8 mL per tube. Purity of the desired peptide in each fraction across the major peak (based on absorbance at 218 nm) was confirmed by electrospray ionization mass spectrometry. Peptides with a purity of 95% or higher were pooled, lyophilized for at least 4 hours (until completely desiccated), and stored at -20°C until disaggregated in preparation for aggregation reactions.

### **2.1.2 Peptide Disaggregation and Reaction Preparation**

To ensure all aggregation reactions were initiated free of any aggregate material that might affect results, all peptides were disaggregated as a standard part of reaction preparation [107, 125, 126]. Pure, lyophilized peptides were brought to room temperature and dissolved in a 1:1 mixture of TFA and hexafluoroisopropanol (HFIP) (Acros Organics) to a peptide concentration of ~100 µg/mL in a glass scintillation vial. The peptide-solvent mixture was left to sit overnight (approximately 16 hours) before evaporation under a stream of nitrogen using the Organomation N-EVAP evaporator (Fisher Scientific). The resulting film was put under vacuum for approximately 1 hour to remove any remaining organic solvent that may have been left behind after nitrogen evaporation. The dried film was then dissolved in acidified water (brought to pH 3.0 with TFA) and centrifuged at 100,000 rpm (416,640 x g) in a bench top ultracentrifuge (Optima TLX, Beckman) for 2-3 hours. Approximately 60-70% of the solution from the top of the supernatant was used as the peptide stock for reaction preparation.

Peptide stock concentration was determined (see below) and subsequently reduced to desired reaction starting values at 1x PBS using 10x PBS (Fisher Scientific) and deionized water. Reaction mixes were filtered with a 20 nm membrane filter (Anotop 10, Whatman) to maximize removal of aggregates. The filtrate was then used as the reaction mixture.

Reactions were conducted at 37°C unless otherwise indicated. Some fibril samples were prepared at low temperatures as described in [127]. In these cases, the prepared reaction mixture was flash frozen and stored at -80°C before transfer to -20°C for incubation.

### **2.1.3 Peptide mass confirmation by mass spectrometry**

The mass of peptides were confirmed with electrospray ionization mass spectrometry on an Agilent electrospray 1100 mass spectrometer. Peptides were injected by auto-sampler onto an Agilent C8 reverse phase column prior to their injection into the MS. Peptide masses were then determined using the in-built Agilent ChemStation software.

### **2.1.4 Peptide concentration determination**

Peptide concentrations were determined using an Agilent analytical RP-HPLC system. Peptides were injected by the auto-sampler onto an Agilent C8 column and eluted using an acetonitrile buffer prepared identically to the purification buffer B. The elution profile was monitored by absorbance at the  $A_{215\text{nm}}$  detector channel and the area of the elution peak was determined using Agilent ChemStation software. The area under the curve was converted to peptide concentration using a standard curve that related amount of peptide to HPLC area [128]. This curve is determined by using the extinction coefficient at 215 nm (calculated using the method described

in [129]) to calculate peptide concentration for 4-5 samples using an absorbance spectrophotometer. These calibrated samples were then injected into RP-HPLC and the area under the curve was plotted against the corresponding concentration determined by the absorbance value to form the standard curve. Generally, each chemically different peptide requires a different standard curve.

### **2.1.5 Dissociation reactions**

Dissociation reactions were initiated with end stage fibril samples. An aliquot was removed from the completed reaction mixture, and diluted in 1x PBS. The goal was to bring the remaining monomer concentration sufficiently below the apparent  $C_r$  value (based on the end-stage plateau monomer concentration from the forward reaction) to allow observation of an increase in monomer concentration as aggregates dissociated when the mixture returns to equilibrium. At the same time, the total amount of peptide in the diluted reaction must be substantially greater than the  $C_r$ , sufficient to ensure that some fibrils will still be present at equilibrium. For example, if a forward reaction had a starting concentration of 100  $\mu\text{M}$ , and appears to equilibrate at 1.5  $\mu\text{M}$  of monomer, a 10-fold dilution will bring the monomer concentration to 0.15  $\mu\text{M}$  (well below the expected  $C_r$ ), while maintaining a total peptide concentration of 10  $\mu\text{M}$  in the reaction mixture, ensuring a clear store of aggregates at equilibrium.

## **2.2 ANALYSIS OF AGGREGATION KINETICS**

### **2.2.1 Sedimentation Assay**

The sedimentation assay protocol is a long-established method in our lab for monomer concentration determination during aggregation reactions [65, 125]. In this method, aliquots of reaction mixture are removed, and centrifuged at 14,000 rpm on a bench top centrifuge (Eppendorf). The top portion of the supernatant (no more than half of the centrifuged volume) was then carefully removed and mixed in a 1:1 ratio with formic acid to prevent further aggregation. The monomer concentration of the reaction (equivalent to the concentration of the monomer remaining in the supernatant) was then determined by injection into a reverse phase HPLC as described in section 2.1.4, above.

### **2.2.2 Thioflavin T Binding**

The ability to bind dyes such as ThT and alter its fluorescence is an identifying characteristic of amyloid fibrils [3]. Though the nature of this interaction is still unclear, this property is commonly utilized to confirm peptide aggregation into amyloid, and to follow aggregation kinetics. While ThT is an excellent test for overall aggregation kinetics and final amyloid formation of polyQ peptides, this method is limited in its ability to report on initial polyQ aggregation kinetics, in particular in a manner sufficiently quantitative to be used in subsequent thermodynamic calculations, and is therefore used sparingly in these studies.

Experiments with ThT binding to polyQ peptides were conducted by mixing an aliquot from aggregation reactions (260  $\mu$ L) with ThT in excess (12  $\mu$ L of 2.5 mM stock). Fluorescence

was measured on a FluoroMax-4 fluorometer from HORIBA Scientific, with an excitation wavelength of 445 nm, a detection wavelength of 489 nm, and slit widths of 2 nm. For all sample measurements, a blank reading of ThT at an identical concentration in 1x PBS free of peptide was subtracted from sample readings.

ThT readings were most frequently reported as a % aggregate by equating the final ThT reading (a plateau) to the final % aggregate as measured by HPLC *via* the sedimentation assay.

### **2.2.3 Dynamic light scattering**

Dynamic light scattering (DLS) was used in these studies primarily to confirm the size of aggregate seeds used for seeding elongation and growing end reactions [18]. For these measurements, 70  $\mu$ L of filtered PBS was added to a 384-well microplate, and a blank reading was taken using a DynaPro plate reader (Wyatt Technology) to confirm the absence of air bubbles and other small particles that might affect a reading of peptide sample. An aliquot of 10  $\mu$ L of the desired peptide sample was then gently mixed into this same microplate well and a measurement was taken immediately to ensure accurate reading of the desired aggregate sample. All measurements were taken at 37°C and for 5 seconds/scan. Typically 10 scans were averaged to determine aggregate size. In all cases, the scans were confirmed to converge to a particular value to ensure accurate reporting of sample size free of contaminants such as micro-bubbles or dust.

#### 2.2.4 Seeded elongation reactions

In all cases, seeded elongation reactions were conducted with end stage aggregates of the desired seed [125]. The weight of peptide incorporated into aggregates was confirmed by subtracting the final monomer concentration from the initial reaction concentration (both as determined by sedimentation assay), this concentration value corresponds to the amount of peptide incorporated into aggregates and were easily be converted to weight per unit volume. Unless specifically noted, seeds for elongation reactions were sonicated on ice using a Fisher Scientific probe sonicator (Sonic Dismembrator, Model 500) at 40% power for a total of 5 minutes, with 30 seconds of sonication alternated with 30 seconds rest. When necessary, seed size was confirmed immediately following sonication by DLS (as described in section 2.2.3).

An aliquot corresponding to a particular amount of aggregate of this sonicated sample was then added to a freshly prepared and filtered reaction. Reaction kinetics were then followed as usual for both this seeded and an identically prepared unseeded reaction.

The initial 20% of the seeded elongation was used to determine the pseudo- first order elongation rate constant,  $k^*$ . These kinetics data were graphed as the  $\ln$  [peptide] versus time, and the slope of the linear fit to the data corresponds to  $k^*$ .  $k^*$  can be considered a pseudo-first order elongation rate constant in seeded reactions where second order nucleation is not a factor because, while the aggregate mass increases, the number of elongation sites does not, thus the elongation rate of this reaction appears to be dependent only on monomer concentration. The actual elongation rate constant for the seeded fibril formation reaction,  $k_+$ , is second order, being dependent both on amount of monomer and the number of elongation sites. For all peptides, two separate seeding reactions (each with seeds prepared from unique samples) were conducted and averaged to determine these values.

### 2.2.5 Growing ends reaction

It is important to determine the number of growing ends present in a reaction in order to calculate the concentration of aggregates and hence the second order elongation rate constant ( $k_+$ ) from  $k^*$ . This second order elongation rate universally applies to all amyloid reactions of a given peptide, and is much more useful toward a complete understanding of peptide aggregation and thermodynamics than the pseudo-first order elongation rate, which has a fibril concentration term embedded within it.

In order to facilitate calculation of  $k_+$ , growing ends titration experiments must be conducted simultaneously with seeded elongation reactions, so that the same aggregate/seed preparation can be used for both the seeded elongation reaction and the growing ends experiment [85, 125]. For growing ends, 200 ng of aggregates are added to a series of reactions containing monomeric biotin-PEG-K<sub>2</sub>Q<sub>30</sub>K<sub>2</sub> at concentrations ranging from 0 to 10  $\mu$ M in a volume of 500  $\mu$ L. The seeds are allowed to incubate with the biotinylated-Q<sub>30</sub> for 30 minutes at 25°C before being centrifuged for an hour at 14,000 rpm and 4°C in a bench top centrifuge (Eppendorf). The supernatant was removed gently, so as not to disturb the pellet, which was then washed in 100  $\mu$ L PBS and centrifuged again at 14,000 rpm for an hour to remove any remaining monomeric biotinyl-Q<sub>30</sub>. This washing step was repeated once more. In all cases, when the supernatant was removed 10% of the volume was left to ensure an undisturbed pellet.

The pellets were then resuspended with 100  $\mu$ L of europium bound streptavidin solution (Perkin Elmer) in a 1:1000 dilution and incubated in the dark for 1 hour at room temperature. The mixtures were again centrifuged and washed twice as previously. The remaining labeled pellets were then mixed with 100  $\mu$ L of enhancer solution (Perkin Elmer) to release europium and the mixtures were then added to a 96 well plate and fluorescence counts were read on a

PerkinElmer/EG&G Wallac Victor<sup>2</sup> microtiter plate reader, which is specially designed to be compatible with the DELFIA/time-resolved fluorescence assays used here. When suitably complexed, europium fluorescence can be used to measure concentrations as low as  $10^{-13}$  M [130]. When measuring concentrations expected to be on the order of fM, such a sensitive assay is not only helpful, but necessary.

Background signal was determined from the preparation of aggregates incubated with 0  $\mu$ M biotinyl-Q<sub>30</sub> as a control. This signal was subtracted from the counts of the other growing ends reaction to determine normalized values. These normalized fluorescence counts were converted to femtomoles of europium using a standard curve determined in-lab from a stock europium solution. Femtomoles europium were then converted to femtomoles biotinyl-Q<sub>30</sub> using the manufacturer's quoted number of europium atoms per streptavidin molecule and an assumed 1:1 binding ratio of streptavidin to biotin molecule. This value was converted to the concentration of growing ends (femtomoles/ $\mu$ g aggregate) based on the use of 200 ng of aggregate per growing ends reaction.

When plotted against the concentration of biotinyl-Q<sub>30</sub> used in each reaction preparation, the number of growing ends is expected to reach a plateau value. This indicates that all growing ends have been saturated, or bound and labeled with biotinyl-Q<sub>30</sub> and Eu-streptavidin, and that this value is the true concentration of growing ends for a given aggregate sample. The validity of the growing ends titration and they underlying basis for how it works are addressed experimentally in Chapter 3.0

## 2.2.6 Aggregation and nucleation kinetics

The Ferrone model [22, 74, 117] discussed in section 1.3.5.1 relates reaction kinetics to nucleation as follows:

$$C(t) = \frac{1}{2} K_{n^*} k_+^2 [C_0]^{n^*+2} t^2 .$$

The variables  $C(t)$  and  $C_0$  refer to the monomer concentration over time and the starting monomer concentration, respectively,  $K_{n^*}$  refers to the nucleation equilibrium constant,  $n^*$  is the nucleus size, and, as mentioned previously,  $k_+$  is the second order elongation rate constant, as described previously. Using measurements of reaction kinetics, which provide monomer concentration with respect to time, we are able to extract the term  $\frac{1}{2} K_{n^*} k_+^2 [C_0]^{n^*+2}$  as the slope from the linear fit to the plot of  $C(t)$  versus  $t^2$ . Graphing the log of this rate versus the log of  $C_0$  allows us to determine values for  $n^*$  (the slope of this graph is  $n^*+2$ ) and  $\frac{1}{2} K_{n^*} k_+^2$  (the y-intercept).

In order to determine  $K_{n^*}$  and  $k_+$  individually, we must determine the concentration of growing ends of the fibrils and the pseudo-first order rate constant, as described above in section 2.2.5. By definition, the pseudo-first order rate constant ( $k^*$ ) is related to the rate constant of a second order reaction ( $k_+$ ) according to the following equation:

$$Rate = k_+ [A][B] = k^* [A].$$

This relation is true when the concentration of B remains constant during the course of the reaction. When an aggregation reaction is seeded with a small concentration of pre-formed fibrils, the reaction proceeds according to first order elongation kinetics and  $k^*$  can be determined as the slope of the line fit to a  $\log(C_0)$  versus time plot. The concentration of growing ends in a given reaction,  $[B]$ , is determined as described **above**. The second order rate constant ( $k_+$ ) can then be determined using a simplification of the above equation:

$$k_+[B]=k^*.$$

Using this value of  $k_+$ , it is now simple to determine the  $K_{n^*}$  of the peptide in question based on the y-intercept of the log-log plot described previously.

While these rates and the nucleus size are essential to our understanding of the nucleation and elongation of amyloid fibrils, it is the change in free energies of nucleation and elongation ( $\Delta\Delta G_{n^*}$ ,  $\Delta\Delta G_{\text{elong}}$ ) that truly allow us to compare the relative impacts of any mutations or variation in these peptides. By definition,

$$\Delta G = -RT \ln K_{eq},$$

and therefore

$$\Delta\Delta G = -RT \ln (K_{eq1}/K_{eq2}).$$

In the above paragraph, we explain how  $K_{n^*}$  is derived: this value is used as the equilibrium constant to determine  $\Delta G_{n^*}$  and  $\Delta\Delta G_{n^*}$ .

The concentration of growing ends in a reaction does not change as the fibrils elongate or dissociate near the equilibrium position. At this point, the reaction is in a state of dynamic equilibrium defined by an equilibrium constant that is the ratio of the elongation rate to the dissociation rate. This term is the binding constant of monomers to fibrils,  $K_a$ , which can be simplified from  $[fibril\ ends]/[monomer][fibril\ ends]$  to simply  $[monomer]^{-1}$  [37]. At equilibrium, this corresponds to the reciprocal  $C_r$ . We can then use the  $C_r$  as the  $K_{eq}$  to determine the  $\Delta G_{\text{elong}}$  and  $\Delta\Delta G_{\text{elong}}$ , and relate stabilities of different fibrils.

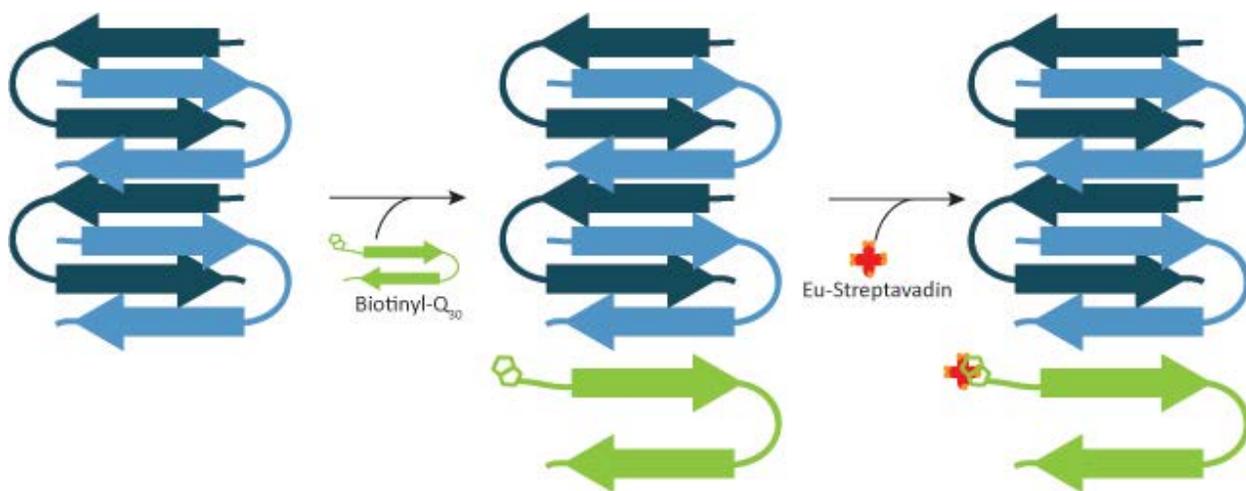
### **3.0 METHOD DEVELOPMENT: UNDERSTANDING THE GROWING ENDS TITRATION EXPERIMENT**

#### **3.1 OVERVIEW**

When studying amyloid growth mechanisms, it is important to understand the rates and equilibria that govern them. For example, the rate at which disordered monomer adds to a growing fibril, the second order elongation rate constant ( $k_+$ ), can be used to identify a number of characteristic features of a particular amyloid sample. This value defines the ability of a particular fibril to seed growth of a range of peptides, and can be used to compare cross-seeding efficiencies of fibrils for a particular monomer [131]. This term is also an essential component of thermodynamic calculations regarding amyloid nucleus stability [22, 132], and is necessary to isolate the value of the nucleation equilibrium constant, which is particularly difficult to measure directly.

Fibril elongation reactions can easily be fit to pseudo-first order kinetics, but the resulting rate constant,  $k^*$ , is specific to the particular seed preparation and concentrations used as a template for elongation, and cannot be generalized to indicate any universal characteristics of fibril elongation for that peptide. Determination of the value of  $k_+$ , in contrast, is much more informative, as it describes the dependence of rate on both monomer concentration and the average number of growth/elongation sites per unit weight of aggregate ([growing ends]) in a

given reaction preparation. While monomer concentration is a fairly straightforward measurement, [growing ends] in a given reaction is difficult to isolate. One approach to determine [growing ends] is to use an electron micrograph of a fibril sample to guesstimate the number of available elongation sites in a given fibril sample [85, 133]. This particular method involves a number of assumptions, such as the idea that each filament terminus is capable of elongating, and that the EM grid is representative of the sample in solution. This approach is clearly less than optimal for use in rigorous calculations that require accurate determination of [growing ends], such as the calculation of  $k_+$  and subsequent determination of  $K_n^*$ .



**Figure 3-1. Growing ends assay.**

Fibril seeds (blue) are incubated with biotinyl-polyQ, which elongates the fibril. The biotin is then labeled with Eu-streptavidin, and Eu fluorescence is measured after incubation with enhancer solution (which frees the Eu molecules from the streptavidin and allows for detection of Eu fluorescence). Not to scale, the streptavidin tetramer is actually much larger than a single polyQ monomer.

A more quantitative experimental approach to measuring [growing ends] has been developed which incorporates labeled monomer onto the ends of growing fibrils *via* a titration assay. The labeled monomer is capable of binding fluorescent molecules, and the resulting fluorescence counts correspond directly to [growing ends] (Method described in 2.2.5) [85, 125].

This assay works through the addition of biotinyl-polyQ (biotin-PEG-K<sub>2</sub>Q<sub>30</sub>K<sub>2</sub>) to existing fibrils, followed by subsequent labeling with europium complexed streptavidin and detection of fluorescence signal (**Figure 3-1**). The efficacy of this method was confirmed by comparing the measured values of [growing ends] to those approximated by the EM method described above [85].

As initially described [85], an important part of the method was the addition of unlabeled K<sub>2</sub>Q<sub>30</sub>K<sub>2</sub> monomer to reactions after the biotinyl-polyQ incubation, in order to induce subsequent rounds of elongation and thereby suppress or prevent dissociation of the labeled molecule and loss of signal. In support of this requirement, when the unlabeled monomer incubation step was skipped, the signal corresponding to binding of the label was lost, suggesting that added biotinyl-polyQ molecules were only weakly and transiently bound to the fibril ends. However, subsequent experience with the assay led to the realization that the addition of unlabeled monomer is actually not required and does not change the signal (MJ, personal correspondence). While the reason for this discrepancy is not clear, it is now routine in the lab to conduct these titrations without an unlabeled monomer “chase”, and excellent dose and time dependent binding is observed [18, 96, 125, 126].

While this method appeared to be a reliable approach to determine [growing ends], the means by which only one biotinyl-polyQ molecule seemed to add to each growing end were a mystery. Our knowledge of elongation rates indicates that polyQ monomers add to fibrils at rates on the order of  $10^4 \text{ M}^{-1}\text{s}^{-1}$ , which corresponds to the addition of 1 molecule every 10 seconds in a 10  $\mu\text{M}$  reaction. The measured addition of only one biotinyl-polyQ to each fibril growing end (with an incubation time of 30 minutes) regardless of whether efforts were made in controlling the reaction conditions to only encourage addition of one biotinyl-polyQ, was surprising, and led

to the hypothesis that the rate of biotinyl-polyQ elongation might be drastically slowed compared to that of simple polyQ. One proposed explanation for this apparent change in rate was based on the “dock and lock” model for fibril elongation—biotinyl-polyQ was capable of undergoing the “dock” phase as efficiently as simple polyQ, but the conformational rearrangement required during the “lock” phase of the mechanism was slowed, delaying presentation of a new elongation site to solution, and decreasing the effective elongation rate of biotinyl-polyQ to fibrils. It is also possible that once bound (“docked and locked”) the biotinyl-polyQ somehow inhibits addition of other monomers. There is currently no experimental evidence for or against either of these arguments, leaving us to wonder why this method appears to function so well.

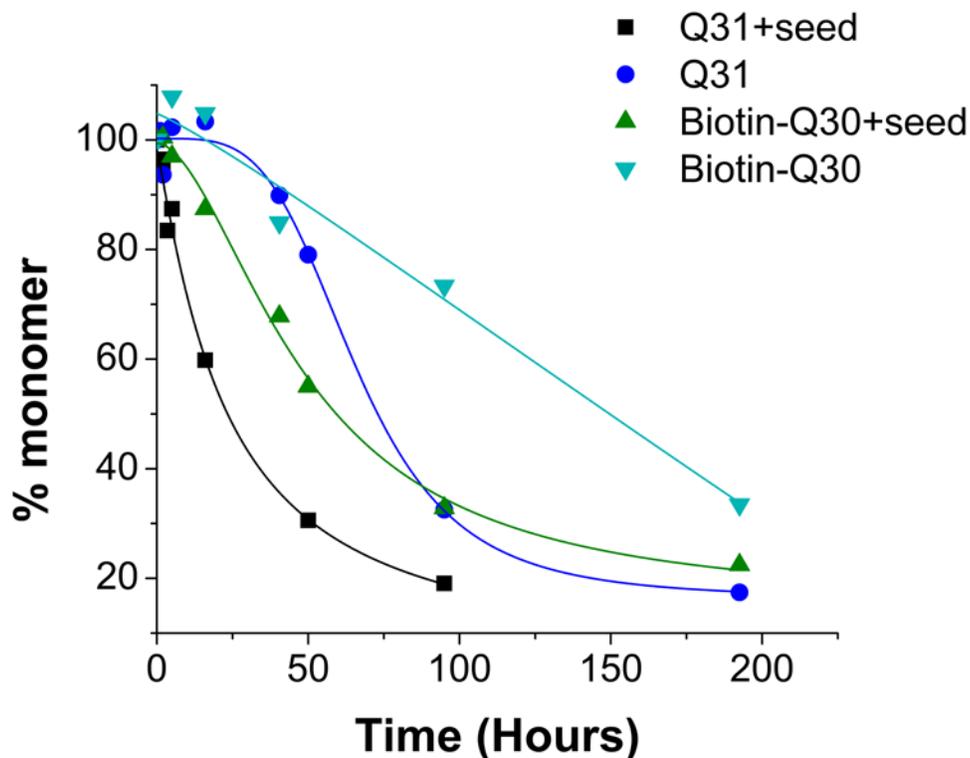
The experiments described in this chapter were conducted in an effort to learn more about this growing ends titration assay, in the hopes that we might be able to arrive at a conclusive mechanism by which it acts as well as to confirm its functionality, which was so surprising when it was first developed.

## **3.2 RESULTS**

### **3.2.1 Biotinyl PolyQ Elongation of PolyQ Fibrils**

In order for biotinyl-polyQ to allow us to accurately measure [growing ends], it must bind and label functional fibril elongation sites in a manner consistent with continuing elongation. If the biotinyl-polyQ were adding to seeds in some unproductive way in the growing ends titration, then one would expect that it would not be capable of multiple rounds of seeded elongation. Seeding reactions were therefore conducted to confirm that biotin-PEG-K<sub>2</sub>Q<sub>30</sub>K<sub>2</sub> is incorporated

into fibrils and elongates seeded material in a manner similar to unlabeled polyQ. **Figure 3-2** illustrates the aggregation reactions of  $K_2Q_{31}K_2$  and biotin-PEG- $K_2Q_{30}K_2$  both in the presence and absence of  $Q_{31}$  seeds. Both peptides proved capable of spontaneous aggregation and demonstrated enhanced aggregation kinetics when incubated with  $Q_{31}$  seeds.



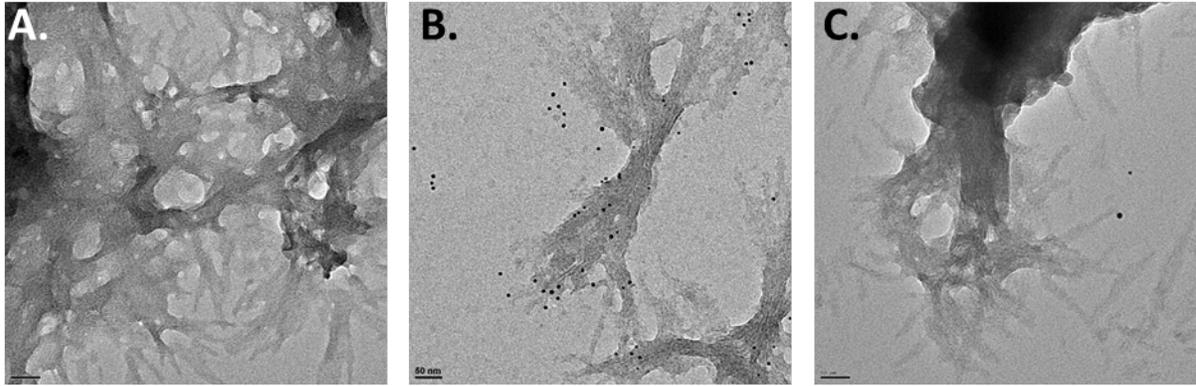
**Figure 3-2. Biotinyl polyQ seeding reactions.**

The aggregation kinetics of  $K_2Q_{31}K_2$  and biotin-PEG- $K_2Q_{30}K_2$  both with and without seed. Concentrations ranged from 22-28  $\mu\text{M}$ , and seeding reactions contained approximately 10% seed (end stage  $Q_{31}$  fibrils with 2.5  $\mu\text{M}$  peptide incorporated into the fibrils). Seeding enhances aggregation for both  $Q_{31}$  and biotinyl-polyQ, and both peptides undergo spontaneous aggregation.

The seeded simple  $Q_{31}$  peptide reached  $t_{1/2}$  (the time at which 50% of monomer is incorporated into aggregates) nearly twice as quickly as biotinyl- $Q_{30}$ , even though the same amount of the same seed was used and the starting monomer concentrations were similar. Biotinyl- $Q_{30}$  also demonstrates slower unseeded (spontaneous) aggregation kinetics when

compared to simple Q<sub>31</sub>. Though it aggregates more slowly than simple Q<sub>31</sub>, Biotinyl-Q<sub>30</sub> does not demonstrate the characteristic lag phase in its spontaneous aggregation kinetics (this lag phase is described in 1.3.1). This could be due to an alteration in the Biotinyl-Q<sub>30</sub> aggregation mechanism—perhaps the biotins associate to encourage formation of oligomers that are pelletable, resulting in this early drop in monomer concentration. If they behave in a manner similar to exon 1 oligomers (described in 1.3.3.2), they could serve to aid amyloid nucleation by creating a local high concentration of polyQ molecules, then dissociate as fibrils form, leaving only monomer and fibril remaining in the end stage reaction. In spite of the slight differences in kinetics, these data confirm that biotinyl-polyQ is capable of spontaneous aggregation and incorporating itself into existing fibrils, suggesting that there is no impediment to the addition of multiple biotinyl-polyQ molecules to the elongation site of a fibril, as previously hypothesized.

To verify that the enhanced aggregation kinetics biotinyl-polyQ seeded reactions were due to elongation of fibrils rather than random association of labeled peptide with seeds, EM was used to observe the morphology of these fibrils (**Figure 3-3**). EM micrographs indicate that biotinyl-polyQ forms amyloid fibrils, both with and without seed present, rather than the disordered aggregates that might be observed if the peptide were incorporated into seeds more randomly. Interestingly, when fibrils elongated with biotinyl-polyQ are fixed to an EM grid and labeled with gold nanobeads coupled to streptavidin, the beads tend to localize preferentially to the termini of fibrils, even when the fibrils are predominantly made up of biotinyl-polyQ.



**Figure 3-3. EM micrographs of biotinyl-polyQ fibrils.**

Biotinyl-polyQ is incorporated into fibrils when seeded with  $K_2Q_{31}K_2$  (A). The morphology is identical for unseeded biotinyl-polyQ fibrils. When labeled with streptavidin-nanogold, the nanogold primarily localizes to the termini of fibrils, even when they contain more than 80% biotinyl-polyQ (B). To confirm that this nanogold interaction is due to the streptavidin attached to the nanogold binding with biotin incorporated into fibrils, the streptavidin-nanogold was bound with free biotin before fibril labeling. No binding is observed for these nanogold particles (C).

These kinetics and EM data indicate that biotinyl-polyQ is capable of adding more than one molecule to the growing end of fibrils. This is surprising because the data used to develop this assay were interpreted to suggest that only one biotinyl-polyQ added per growing end [85]. This result introduces a series of important questions: if there is incorporation of more than one labeled molecule onto the fibril termini, why is it that the number of bound biotinyl-polyQs plateaus after a short incubation time, and why does this value correspond so well to the estimated number of growing ends in a sample? The nano-gold labeled EM samples suggest a possible new mechanism: only the biotin molecules on the termini of the fibrils remain accessible.

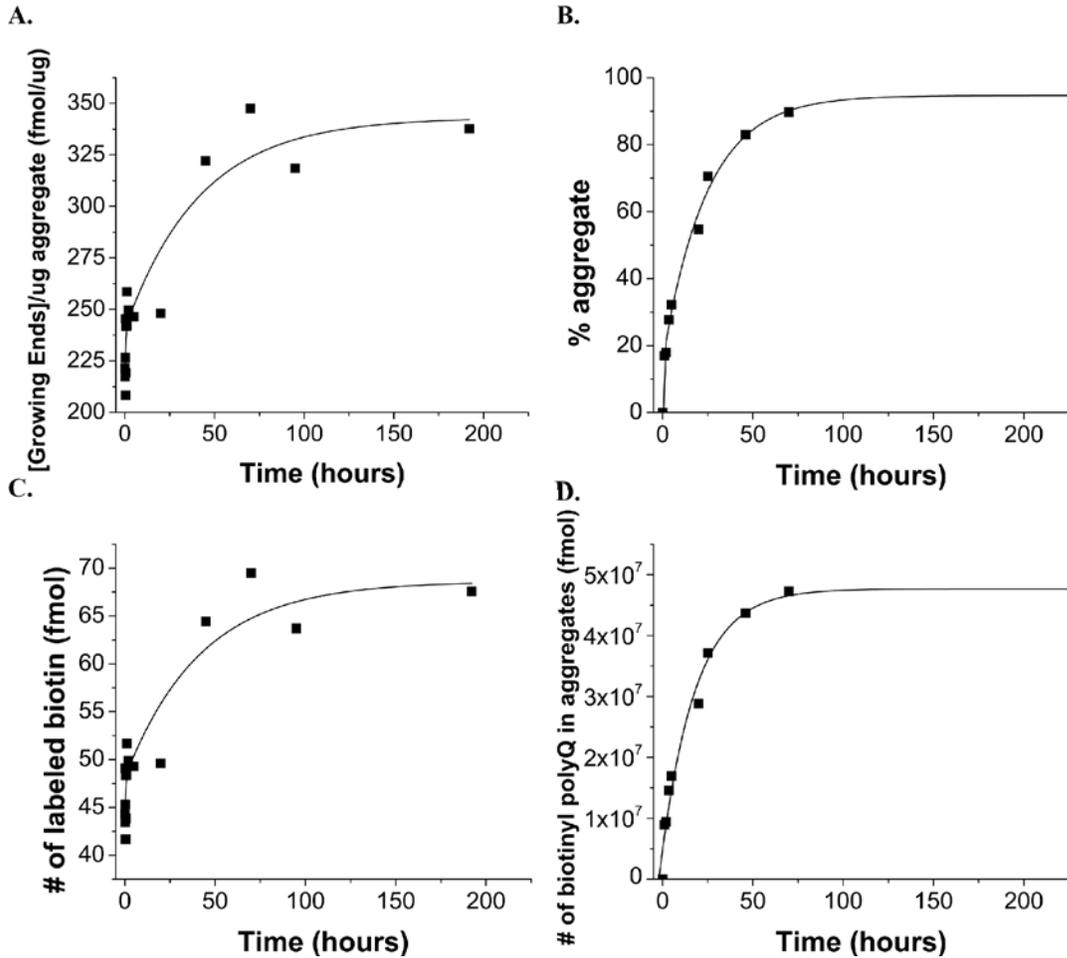
### 3.2.2 Time dependent elongation with biotinyl-polyQ

Time dependent fluorescence studies of biotin-Q<sub>30</sub> incorporation into fibril seeds similar to those conducted during original development of this method [85] were repeated, this time on longer time scales (hours as compared to minutes). While the fluorescence counts in these studies did increase at a rate similar to the observed rate of aggregation (as measured by sedimentation assay), they are not consistent with the amount (in moles) of biotinyl-polyQ that have been incorporated into fibrils (**Figure 3-4**). In fact, the amount of incorporated biotin as detected by fluorescent labeling is roughly 5 orders of magnitude lower than the amount indicated by sedimentation assay!

The inefficient labeling with Eu-streptavidin implicit in **Figure 3-4 C, D** is consistent with the streptavidin-nanogold labeled EM results, in which fibrils appear to selectively bind nanogold at the termini. Considered together, these EM and time-course fluorescence data suggest that biotinyl-polyQ incorporated in the interior of the fibril sequester the biotin molecule preventing binding with streptavidin in solution. That is, biotinyl-polyQ molecules that have bound the growing end of the fibril maintain a biotin-available state until a conformational change or addition of the next monomer sequesters it, preventing binding of streptavidin bound nanogold or Eu (see mechanism in **Figure 3-9**).

Fibril formation and elongation are reversible folding and association reactions, in which there is constant ‘recycling’ of molecules in the system as those at the termini interchange with those in solution [134]. Previously, unlabeled monomer was introduced to the reaction subsequent to the labeling reaction to prevent dissociation of the labeled polyQ *via* this molecular recycling [85]. However, if multiple labeled molecules are incorporated into the fibrils but only the ‘last on’ is counted, this incubation with unlabeled monomer step is unnecessary, as use of a high enough

concentration of biotinyl-polyQ to add multiple molecules during the labeling portion of the assay plays a similar role—one accessible biotin molecule is always maintained for each growing end, even if those near the termini are cycling with monomers in



**Figure 3-4. Time dependence of biotinyl-polyQ incorporation.**

The elongation kinetics of 5  $\mu$ M biotin-PEG- $K_2Q_{31}K_2$  with 5%  $K_2Q_{25}K_2$  -20°C fibril seed (monofilaments) as observed by fluorescence (A, C) and a corresponding (scaled up) reaction followed by sedimentation assay (50  $\mu$ M start concentration, with 5% seed) (B, D). Both approaches demonstrate similar kinetics, but the plateau level of the fluorescence reading does not correspond to the amount of biotin that has been incorporated into fibrils (C vs. D)—when 89% of biotinyl-polyQ are incorporated according to sedimentation assay kinetics, the Eu signal maximum has less than doubled and is consistent with available biotin equivalent to only 0.003% of the biotinyl-polyQ.

solution. This incubation with unlabeled monomer step has been confirmed to be unnecessary within our lab (MJ, personal correspondence), a conclusion is further corroborated by the fluorescence data shown in this chapter, which shows significant Eu signal even without the addition of unlabeled monomer.

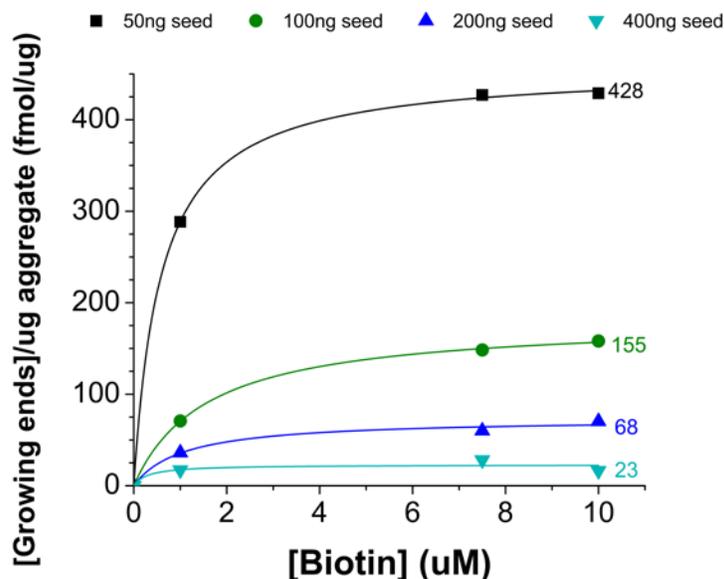
In addition, if we use the data illustrated in **Figure 3-4** to approximate the number of elongating biotinyl-polyQ molecules that are incorporated into each growing end, we find that there are nearly 35 thousand biotin molecules adding to each growing end (assuming equivalent elongation preference for each end). Thus, even if the terminal-most thousand biotinyl-polyQ dissociate at some point during the work up of this assay, there are still thousands more incorporated into the fibril capable of accepting a streptavidin bound label.

### **3.2.3 Growing ends titration accurately reports available growing ends**

Further confirmation of method functionality is necessary to be certain it accurately reports the number of growing ends. If the method functions as proposed, an increase in signal should correspond to a similar increase in number of available growing ends. By using a seeded elongation coupled with the growing ends titration assay on seed material known to vary in the number of available elongation sites, it should be possible to confirm method functionality. There are a number of approaches that can be used to generate variable numbers of elongation sites in otherwise identical fibril seed samples, but the two most straightforward approaches are by varying seed mass used in the growing end titration, and by maintaining the same seed mass but fragmenting the fibrils to result in variable numbers of growing ends.

### 3.2.3.1 A first attempt: varying the seed mass

The simplest approach to introducing a variable number of elongation sites to the biotin titration assay would seem to be using different amounts (by weight) of seed in the assay. Usually 200 ng of seed are used for the titration assay, so experiments were conducted using both more (400 ng) and less (100 and 50 ng) of an identical seed preparation, while maintaining all other conditions equal. Using this approach, the titration assay reactions with more seed should have higher fluorescence counts, corresponding to more growing ends, while those with less seed should have lower fluorescence counts, as they have fewer elongation sites. The data for these experiments are shown in **Figure 3-5**.



**Figure 3-5. Growing ends titration data from reactions with variable amounts of seed.**

Four masses of seed were used to perform the growing ends assay. All four samples resulted in similar fluorescence count values, which, when corrected for the amount of aggregate to generate fmol of growing ends/  $\mu\text{g}$  aggregate, leads to a surprising result. Smaller amounts of seed appear to generate more signal per growing end than larger amounts.

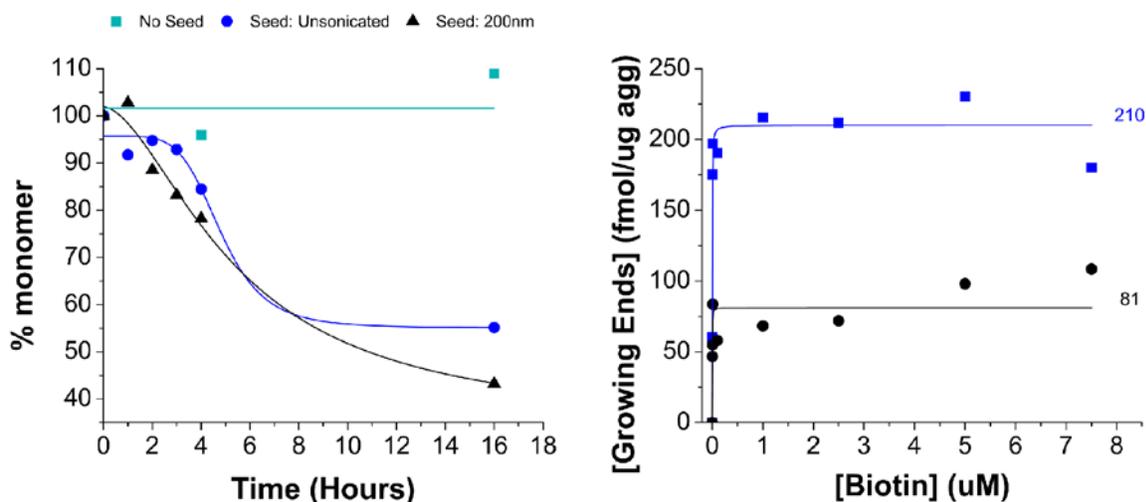
Contrary to expectations, larger amounts of seed appeared to have fewer labeled growing ends while the titrations with less seed gave higher fluorescence readings. A possible explanation

for this lies in the centrifugation steps used to isolate biotinyl-polyQ bound aggregate seeds. Larger amounts of aggregate could result in a greater number of aggregates being buried within the pellet prior to the Eu-streptavidin labeling step. These inaccessible aggregates then are not completely resuspended during Eu-streptavidin labeling, and are therefore not fully labeled. Thus when fluorescence counts are measured, amounts of seed >200 ng appear to have fewer growing ends, even though this is not the case and Eu-streptavidin was added in excess of the possible number of added sites.

Since the use of variable amounts of seeds was not an effective means of confirming that this biotinyl-polyQ titration assay functions as predicted, and in fact gave counterintuitive results, another approach was necessary, this time designed to maintain the same weight of aggregates, but with variable amounts of elongation sites.

### **3.2.3.2 Sonication generates samples with variable amounts of elongation sites**

Sonication of amyloid fibrils has been demonstrated to result in an improvement in elongation efficiency. This change is proposed to correspond with an increase in elongation competent fibril ends that result from fracturing of the fibrils [125]. By applying varying sonication energies and times to fibril samples from the same stock preparation, it should be possible to generate a series of seed samples with variable amounts of growing ends. When the growing ends titration is conducted with these samples, each unique seed sample should yield a number of growing ends consistent with the amount of fracturing it has undergone.

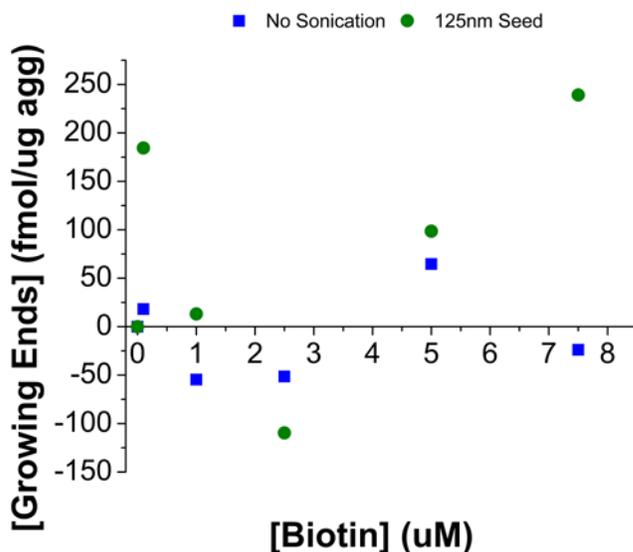


**Figure 3-6. Seeded elongation and growing ends data (14 krpm centrifugation).**

Seeded elongation of Q<sub>25</sub> with 6% seed demonstrates sonication of the aggregate used as seed has a very slight enhancement on elongation ability when compared to use of a larger, unsonicated seed (~2000 nm). Both seeded reactions accelerate significantly more rapidly than an otherwise identical unseeded reaction. The growing ends titration assay, however, indicates that the large, unsonicated seeds have approximately 2.5 times the number of growing ends as sonicated seeds, an unexpected result.

For these experiments, fibrils were prepared using the Q<sub>25</sub> peptide at -20°C. Under these conditions, polyQ forms fibrils that appear to be a single filament by EM, and these more fragile fibrils are capable of being fractured by sonication into seeds of sizes as small as 50 nm (when measured by DLS) [127, 135]. Samples of 200 ng of aggregate were used as seed for this study, consistent with the typical approach for the method (2.2.5). Sonication for 10 sec at a power of 40% with a probe sonicator (2.2.4) generated 200 nm seeds. When unsonicated, fibrils had an average size of ~2000 nm. The growing ends titration assay was followed as described in Methods section 2.2.5 using these two seed samples. Contrary to expected results, smaller aggregates appeared to have fewer elongation sites, in spite of a slightly enhanced seeding ability (**Figure 3-6**). This result suggests that the 14 krpm centrifugation that is used to pellet labeled aggregate may not be generating sufficient force (or may not have been run long enough) to

quantitatively pellet the small, 200 nm seeds. Loss of these seeds during any of the many centrifugation steps could explain the artificially low number of elongation sites we detect as during the growing ends titration experiment. Therefore, a second attempt was made using an ultracentrifuge to pellet the aggregate seeds at 100 krpm for all centrifugation steps. The results of the growing ends titration (**Figure 3-7**) are not interpretable and suggest that an alternative to the centrifuge must be used to isolate labeled aggregates.



**Figure 3-7. Variable seed size growing ends titration (ultracentrifugation).**

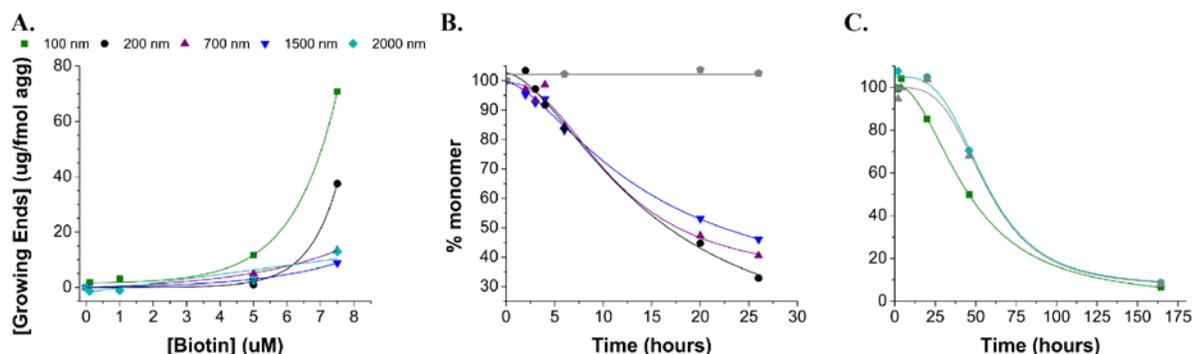
When ultracentrifugation is used to pellet aggregates, there is no detectable trend in labeling of sample or number of elongation sites. The inconsistency of this data proves the ineffectiveness of this approach to confirm functionality of the growing ends titration assay.

**Table 3-1. Sonication impacts on growing ends.**

<b>Seed size (determined by DLS)</b>	<b>[growing ends]/μg aggregate</b>
<b>100 nm *</b>	71
<b>200 nm</b>	38
<b>700 nm</b>	13
<b>~1500 nm (unsonicated)</b>	9
<b>~2000 nm (unsonicated) *</b>	13

Experiments were performed on two separate seed preparations with otherwise identical characteristics, starred values were performed on one sample, while unmarked values a second.

One final attempt was made to generate and study variably sized seeds, the characteristics of which are listed in **Table 3-1**. As an alternative to centrifugation, a filter trap assay was used to recover and count labeled aggregates. Biotinyl-polyQ labeled seeds were prepared as usual, but before incubation with Eu-streptavidin, they were run through a vacuum filtration apparatus with a cellulose acetate filter (Whatman) with a pore size of 200 nm. A series of wash steps were run before and after Eu-streptavidin was added to label the bound fibrils. The membrane containing the labeled aggregate was then submerged in 100  $\mu$ L enhancer solution, and sonicated for 2 min in a bath sonicator to free the bound Eu. All 100  $\mu$ L were then removed to a 96 well plate, and the fluorescence counts were measured as usual.



**Figure 3-8. Seeded elongation and growing ends titration with variably sized seeds (filter trap).**

When labeled fibrils are isolated using a filter trap, [growing ends] increases as seed size is decreased (A). Seeded elongation reactions were conducted simultaneously, with 25% (B) or 0.5% (C) seed. In both cases, smaller seeds are more efficient at elongation. This improvement in elongation is more evident when more seed is used, as in (B).

The resulting amounts of growing ends are listed in **Table 3-1** and shown in **Figure 3-8**. A. These values are consistent with the expected trend based on sonicated seed size, and enhancement of aggregation kinetics observed in corresponding seeded elongation reactions (**Figure 3-8 B and C**). The smallest seeds provide the most growing ends for the same mass of seeds, while those that have undergone no sonication (and are the largest) have the least.

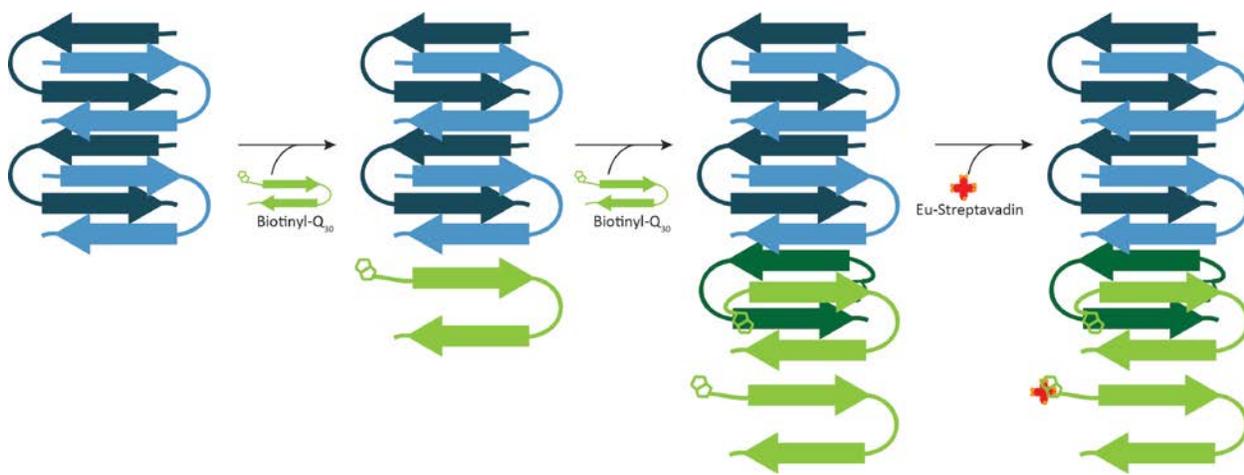
Unfortunately, it is impossible to verify a loss of seeds during the filter trap assay, making a quantitative comparison of these values difficult to justify. It is possible that growing ends values for the small seeds in particular may be artificially low due to loss of aggregates during this step. Since these small seeds still generate the expected higher fluorescence counts when compared to large seeds, we believe any sample loss that may occur is fairly modest, and does not impede our confirmation of the functionality of this mechanism.

### 3.3 DISCUSSION

The goals of the work described in this chapter were to attempt to confirm functionality and to develop a mechanism that could explain how a commonly used lab assay is able to accurately report the number of elongation sites in a fibril sample. The series of experiments conducted to this end led to mixed results which were often difficult to interpret.

Experiments indicating increases in the number of elongation sites in a given sample of seed correspond to increases in the measured number of growing ends (as observed by fluorescence counts) suggest that this assay is capable of at least qualitatively reporting the amount of elongation sites in a given sample. These functionality studies also indicate that the mass of fibrils studied (200 ng) and the average seed size are important to accurate measurement of [growing ends]. If too little or too much aggregate (by mass) is used, there may be effects on the efficiency of the labeling step. Likewise, the use of seeds that are very small (<100 nm), prevents formation of a dense pellet during centrifugation, and leads to inaccurate results. Use of a filter-trap apparatus makes isolation of small aggregates possible, but it is difficult to confirm that no sample is lost using this method.

The data presented here point to an experimental mechanism in which multiple biotinyl-polyQ elongate the fibril seed during the labeling phase. Our ability to match [growing ends] counts determined by multiple methods may be because the terminal-most biotinyl-polyQ maintains a solvent accessible biotin molecule, while those more internal to the fibril have undergone a conformational change during a “locking” phase of elongation, sequestering the biotin molecule, so they are no longer accessible to streptavidin in solution. If this proposed mechanism is correct, the growing ends titration experiment can be expected to consistently report the amount of growing ends in the reaction, so long as spontaneous aggregation of the biotinyl-polyQ does not occur.



**Figure 3-9. Proposed mechanism for growing ends titration assay.**

Incubation of polyQ seeds with biotinyl-polyQ leads to addition of multiple biotinyl-polyQ molecules. When the biotinyl-polyQ first binds/“docks” the biotin remains accessible. Upon “locking” the biotinyl-polyQ undergoes a conformational change to sequester the biotin, making it inaccessible within the fibril. Biotinyl-polyQ on the growing end that has not undergone conformational change maintains an accessible biotin that binds Eu-streptavidin and yields the fluorescence signal.

In the original publication describing this method, experiments in which labeled fibrils were incubated with simple Q<sub>30</sub> in a quenching step still generated signal, indicating that

internalized biotinyl-polyQ are capable of binding streptavidin [85]. There are a number of differences between that experiment and those presented here, however, that could explain these inconsistent results. For one, labeled fibrils were elongated with simple polyQ, not biotinyl-polyQ. It could be that the internal biotin molecules are only sequestered when biotin groups on nearby labeled molecules interact, so elongation with simple polyQ leaves biotin molecules free to bind streptavidin.

It is also possible that the binding ratio of streptavidin to biotin is less than 1:1. A single streptavidin molecule is capable of binding multiple biotins, so if multiple biotin groups are close enough together (as they might be after a short elongation reaction, such as the one conducted in [85]) there might be only one label per growing end, either because streptavidin is bound to multiple biotins, or because it is blocking the binding of other streptavidin. In a reaction in which biotinyl-polyQ elongates a fibril over a long time frame (such as the experiments shown in **Figure 3-4**), we would expect the portion of the fibril containing biotin molecules to ultimately lengthen beyond the reach of a single streptavidin molecule, resulting in an increase in signal as new biotins become accessible. While the data shown in **Figure 3-4** indicate the number of labeled sites does increase beyond 2 hours (the scope of the experiment in [85]), the plateau value still corresponds to less than double the amount of peptide being labeled between 2 and 200 hours, when the sedimentation assay indicates that more than double the amount of biotinyl-polyQ has been incorporated into aggregates. If this effect were solely due to streptavidin molecules binding more than one biotin, 1 streptavidin would have to bind or block approximately  $10^6$  biotin molecules, which is orders of magnitude beyond the two biotin binding sites that actually exist in a streptavidin molecule, and significantly higher than we might expect

to see from a blocking effect. Thus, the above proposed mechanism (**Figure 3-9**) is still the most likely explanation for how this biotin titration assay functions.

Unfortunately, the experiments performed in this chapter do not provide a satisfactory mechanistic basis for this assay. The repeated challenges in acquiring data with straightforward interpretations, and inconsistencies between more recently conducted experiments and those used in the development of this assay force the question of whether this assay is accurately reporting the values it was designed to measure. Data acquired using this approach in Chapters 4.0 and 5.0 are also confusing. Luckily, it is still possible to determine a number of thermodynamic parameters that define the polyQ aggregation mechanism in the absence of this assay; it is these parameters that are the main focus of the next two chapters.

## 4.0 TERMINAL CHARGE STUDIES OF POLYQ PEPTIDES

### 4.1 OVERVIEW

As discussed in the Chapter 1.0, the flanking sequences of polyQ expansions play fundamental roles in disease progression and aggregation mechanism, both *in vivo* and *in vitro*. The use of positively charged, Lys residues as solubilizing agents for *in vitro* studies of simple polyQ aggregation has been brought into question due to the electrostatic impact charged termini may have on the peptide aggregation mechanism [122]. By adding positive charges to the termini of polyQ peptides, it might be argued that stereoelectronic hindrances could be introduced that affect a peptide's ability to assume structures essential to its aggregation, such as the  $\beta$ -hairpin nucleus, or the  $\beta$ -strand-turn-strand complex fundamental to the polyQ amyloid structure.

In an effort to identify both the qualitative and quantitative roles charged flanking sequences have on aggregation mechanism, rates, fibril stability and other fundamental thermodynamic terms, we have designed a series of 10 peptides. These peptides consist of a set of flanking charges in the context of two polyQ core sequences. The first core sequence, containing a central D-Pro-Gly, was selected to gain a quantitative perspective on the rates of aggregation with minimal effects on the mechanism (1.4.1.1). A simple polyQ core, consisting of 26 Gln residues was chosen as a more physiologically relevant option. Since the Gln sequence is rarely interrupted in disease states, it is important to consider a simple polyQ model. The repeat

length of 26 was chosen in an effort to maximize the amount of polypeptide involved in essential structure (i.e. stabilizing  $\beta$ -sheet), assuring maximal interaction of charged terminal residues while maintaining an  $n^*=1$ . Since maintaining a monomeric nucleus is essential for quantitative comparisons, a repeat length of at least 26 was required given our prior knowledge of  $n^*$  dependence on repeat length [18]. This simple polyQ peptide serves as a qualitative measure of the role of charged flanking residues on the aggregation mechanism, as it is more prone to mechanistic change than the D-Pro-Gly containing peptide, which has steric constraints encouraging structure formation.

**Table 4-1. Terminal charge peptides and charge characteristics.**

Peptide	Symbol	Net charge	Repulsion	
			Intramolecular	Intermolecular
<b>K<sub>2</sub>Q<sub>26</sub>K<sub>2</sub></b>	□	3.8	2.8	14.44
<b>KQ<sub>26</sub>K<sub>3</sub></b>	○	3.8	3.6	14.44
<b>Q<sub>26</sub>K</b>	▽	0.8	0	0.64
<b>KQ<sub>26</sub>D</b>	◇	-0.2	-3.6	0.04
<b>K<sub>2</sub>Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K<sub>2</sub></b>	■	3.8	2.8	14.44
<b>KQ<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K<sub>3</sub></b>	●	3.8	3.6	14.44
<b>KQ<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K</b>	▲	1.8	0	3.24
<b>Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K</b>	▼	0.8	0	0.64
<b>KQ<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>D</b>	◆	-0.2	-3.6	0.04
<b>K<sub>2</sub>Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>D<sub>2</sub></b>	◆	-0.2	-8.4	0.04

Symbols are used in figures throughout this chapter to illustrate properties of these peptides unless otherwise indicated. All D-Pro-Gly peptides are indicated by filled symbols, while simple polyQ are indicated by unfilled symbols.

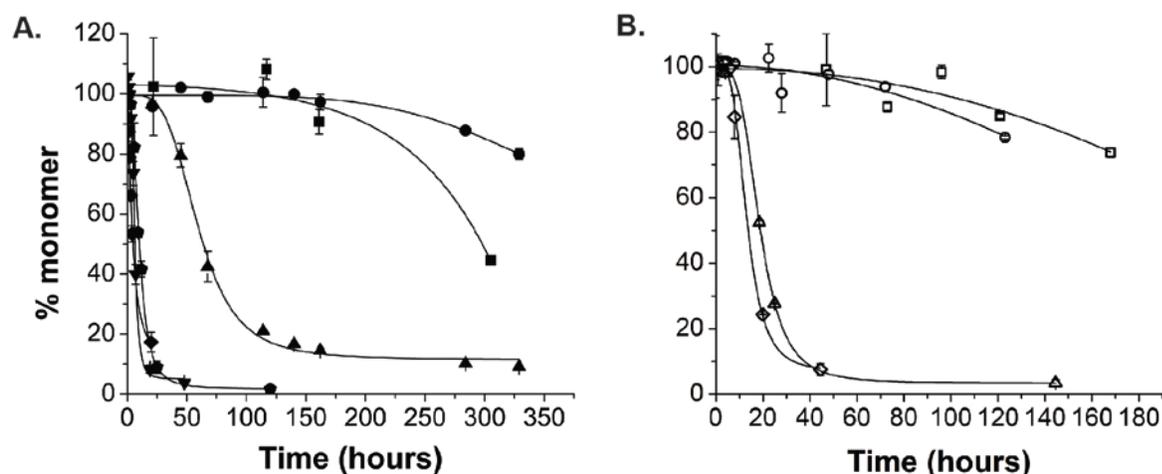
The peptides investigated in this study are listed in **Table 4-1**, along with their charge properties. The net charge of each molecule was calculated by estimating the approximate charge on the peptide at pH 7.3 (as determined using the null model [136], in which an average  $pK_a$

value determined from 78 proteins is used to determine charge [137]), with the N-terminal  $\alpha$ -amino group ( $pK_a \approx 8$ ) assigned a charge of +0.8, Lys side chains ( $pK_a = 10.4$ ), +1, Asp side chains ( $pK_a = 3.9$ ), -1, and C-terminal  $\alpha$ -carboxyl group ( $pK_a \approx 3.7$ ), -1. Repulsion was considered to be proportional to the force experienced between charges as defined by Coulomb's law ( $F = A |q_1 q_2| / r^2$ , where  $F$  corresponds to force,  $A$  is a constant,  $q_{1,2}$  are the two charges in question, and  $r$  is the distance between them). Therefore, the intramolecular repulsion was calculated by multiplying the summed charges on the termini, and intermolecular repulsion was calculated by squaring the net charge of the molecule (identical to multiplying the charge on two like molecules).

## 4.2 RESULTS

### 4.2.1 Impact of Terminal Charges on Kinetics and Morphology

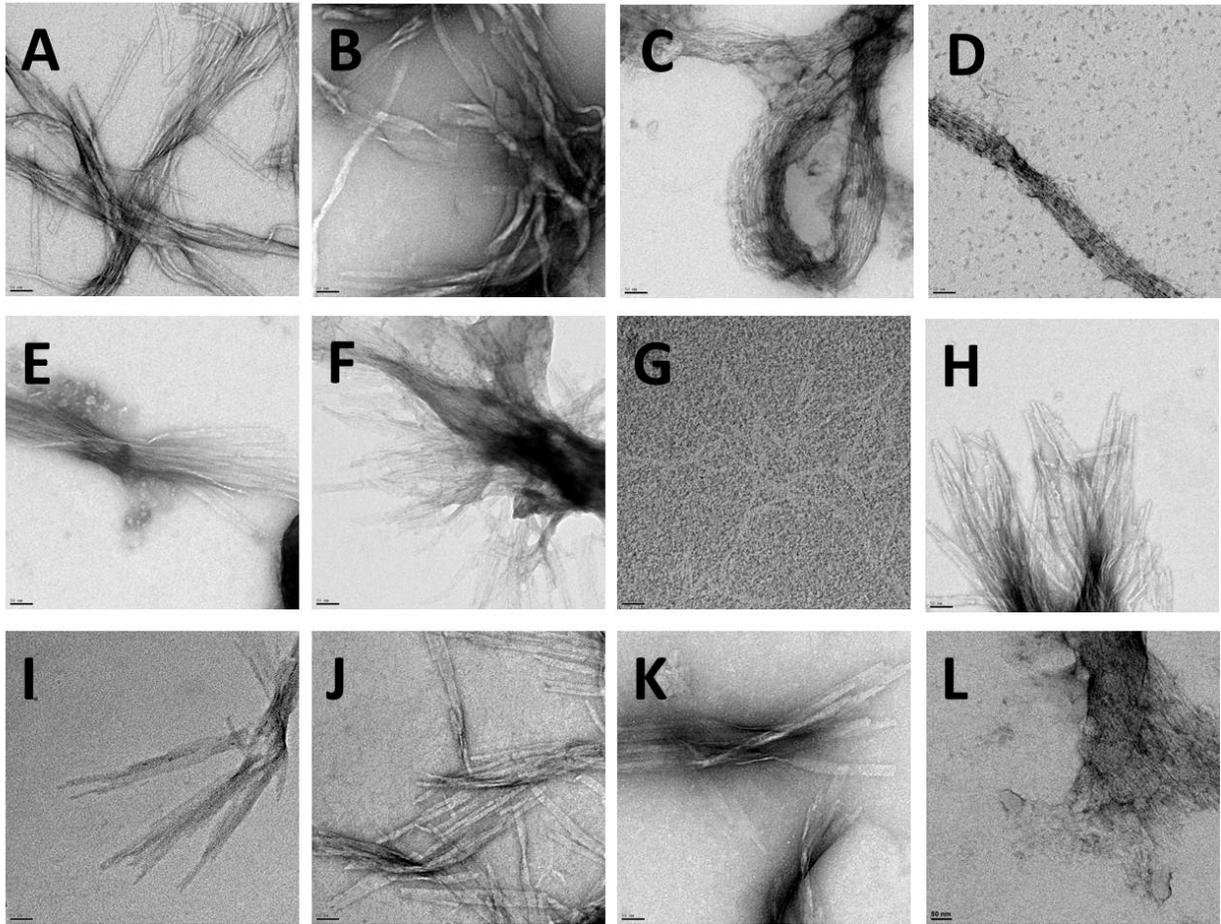
All of the peptides investigated this study proved to be capable of spontaneous amyloid formation. A preliminary survey of the aggregation kinetics of each of the 10 peptides (at similar concentrations) is illustrated in **Figure 4-1**. In general, we found that as repulsion between the peptide termini is decreased, aggregation rate increases in both model frameworks. These changes in aggregation rate are early indications that varying terminal charge does have an impact on the kinetics of polyQ aggregation. Further thermodynamic analysis is necessary, however, in order to determine the specific role these flanking sequences have on the polyQ aggregation mechanism.



**Figure 4-1. Aggregation kinetics of all terminal charge peptides.**

Changes in terminal charge do have a notable effect on aggregation rate, in spite of similar initial monomer concentrations. D-Pro-Gly containing peptides are illustrated in (A) and simple polyQ in (B) at the following start concentrations:  $K_2Q_9P^D GQ_9K_2$  80  $\mu M$  ( $\blacksquare$ ),  $KQ_9P^D GQ_9K_3$  61  $\mu M$  ( $\bullet$ ),  $KQ_9P^D GQ_9K$  69  $\mu M$  ( $\blacktriangle$ ),  $Q_9P^D GQ_9K$  68  $\mu M$  ( $\blacktriangledown$ ),  $KQ_9P^D GQ_9D$  55  $\mu M$  ( $\blacklozenge$ ),  $K_2Q_9P^D GQ_9D_2$  60  $\mu M$  ( $\blacklozenge$ ),  $K_2Q_{26}K_2$  40  $\mu M$  ( $\square$ ),  $KQ_{26}K_3$  34  $\mu M$  ( $\circ$ ),  $Q_{26}K$  40  $\mu M$  ( $\nabla$ ), and  $KQ_{26}D$  40  $\mu M$  ( $\diamond$ ).

End stage aggregate morphology was also confirmed *via* EM (**Figure 4-2**). The morphology of most of the fibrils appears to be within the range of what we typically observe for simple polyQ peptides, but non-amyloid structures were present in the EM images of a few peptides. In these latter cases, the EM micrographs indicated these alternate morphologies were co-populating the reaction with amyloid fibrils. This introduction of new morphologies may be a primary indication of mixed aggregation mechanisms for some peptides in this series.



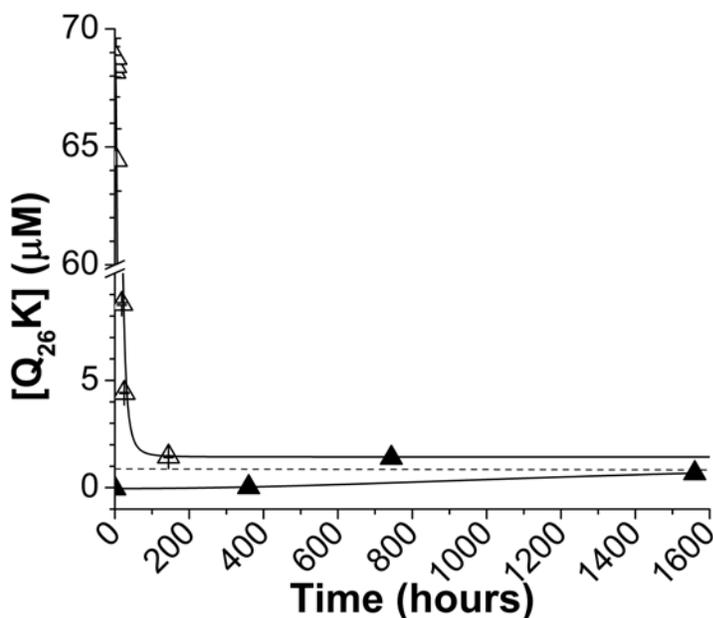
**Figure 4-2. EM images of end stage aggregates.**

Characteristic EM images of all ten peptides: each peptide is capable of adopting the characteristic polyQ fibril morphology, but some generate alternative structures (C, D, G, H, L). The peptides imaged and the reaction start concentration from which they originated are as follows: (A)  $K_2Q_9P^DGGQ_9K_2$ : 181  $\mu$ M, (B)  $KQ_9P^DGGQ_9K_3$ : 45  $\mu$ M, (C)  $KQ_9P^DGGQ_9K$ : 150  $\mu$ M, (D)  $KQ_9P^DGGQ_9K$ : 12  $\mu$ M, (E)  $Q_9P^DGGQ_9K$ : 88  $\mu$ M, (F)  $KQ_9P^DGGQ_9D$ : 78  $\mu$ M, (G)  $K_2Q_9P^DGGQ_9D_2$ : 112  $\mu$ M, (H)  $K_2Q_9P^DGGQ_9D_2$ : 46  $\mu$ M, (I)  $K_2Q_{26}K_2$  43  $\mu$ M, (J)  $KQ_{26}K_3$  63 $\mu$ M, (K)  $Q_{26}K$  108 $\mu$ M, (L)  $KQ_{26}D$  20 $\mu$ M

#### 4.2.2 Fibril Stability

The role of terminal charge electrostatics on fibril stability can be determined using the  $C_r$  (critical concentration) of a given peptide. The  $C_r$  is determined by measuring the final

concentrations of the fibril dissociation and/or association reactions (**Figure 4-3**, 2.1.5). The free energy of fibril elongation is then approximated using the expression:  $\Delta G_{\text{elong}} = -RT \ln (1/C_r)$ . This free energy describes the energy required to add a disordered monomer to the elongating fibril—a more positive value indicates a less stable fibril structure, while a more negative value corresponds to a more stable fibril. For the fibrils investigated in this study, the  $\Delta G_{\text{elong}}$  values lie within a narrow range, between -32 and -36 kJ/mol (**Table 4-2**). This small difference indicates that terminal charges have only subtle effects on fibril stability.

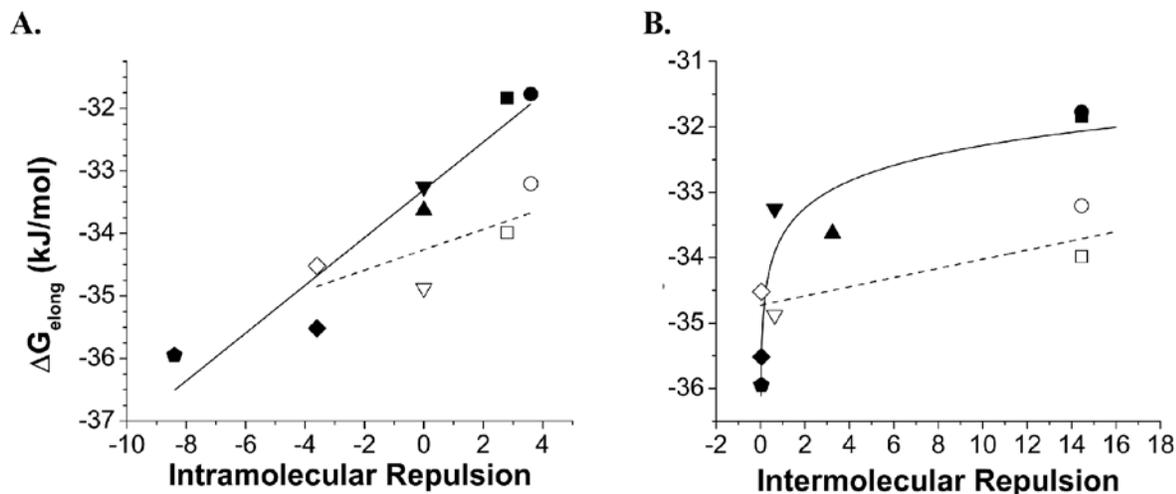


**Figure 4-3.** Association and dissociation reactions meet at the  $C_r$ .

Association (open triangles) and dissociation (filled) reaction of  $Q_{26}K$ . The original reaction start concentration was 68.2  $\mu\text{M}$ . Both association and dissociation approach a  $C_r$  value of 0.66  $\mu\text{M}$ , indicated by the dashed line.

Even within these subtle changes in free energy, it is possible to identify clear influences of flanking charges on fibril stability. For example, a comparison of  $\Delta G_{\text{elong}}$  values with the intramolecular repulsion term (**Table 4-1**) demonstrates that fibril stability increases with the attraction between terminal charges (**Figure 4-4 A**). The most stable fibrils are those with

opposite (attractive) terminal charges ( $\diamond, \blacklozenge, \blacklozenge$ ), while like (repulsive) charges ( $\square, \circ, \nabla, \blacksquare, \bullet, \blacktriangle, \blacktriangledown$ ) lead to a decrease in stability. The linear correlation of  $\Delta G_{\text{elong}}$  with intramolecular repulsion is fairly robust, with the D-Pro-Gly series yielding a slightly steeper slope than the simple Q<sub>26</sub> peptides. These data support the hypothesis that successful and favored closure of a  $\beta$ -hairpin is important to fibril stability and the energetics of amyloid formation.



**Figure 4-4. Free energy of elongation vs. repulsion.**

Free energy of elongation, as calculated from the critical concentrations of all peptides studied graphed against the intramolecular (A) and intermolecular (B) repulsion of each peptide.

The fact that intramolecular repulsion (**Figure 4-4 A**) proves to be more destabilizing to fibril structure in the D-Pro-Gly series than the simple polyQ series justifies our selection of two distinct core structures for these experiments. The gentler slope of the fit to the Q<sub>26</sub> trend indicates that this series experiences relatively little effect in fibril stability as a result of variations in terminal charge, most likely due to flexibility in the termini of these peptides which allows the terminal segments to pull apart without preventing amyloid formation (**Figure 1-7**). Thus, K<sub>2</sub>Q<sub>26</sub>K<sub>2</sub> fibrils ( $\square$ ) are only destabilized by a  $\Delta\Delta G_{\text{elong}}$  of 0.54 kJ/mol compared to KQ<sub>26</sub>D fibrils ( $\diamond$ ), while K<sub>2</sub>Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K<sub>2</sub> fibrils ( $\blacksquare$ ) are destabilized by a  $\Delta\Delta G_{\text{elong}}$  of 3.67 kJ/mol

compared to  $\text{KQ}_9\text{P}^{\text{D}}\text{GQ}_9\text{D}$  (◆) (Table 4-2). The increased impact of terminal charges on the stability  $\beta$ -hairpin mimetic peptide fibrils is predominantly due to our design criterion that minimized chain terminus flexibility in the D-Pro-Gly fibrils.

**Table 4-2. Terminal charge thermodynamic values.**

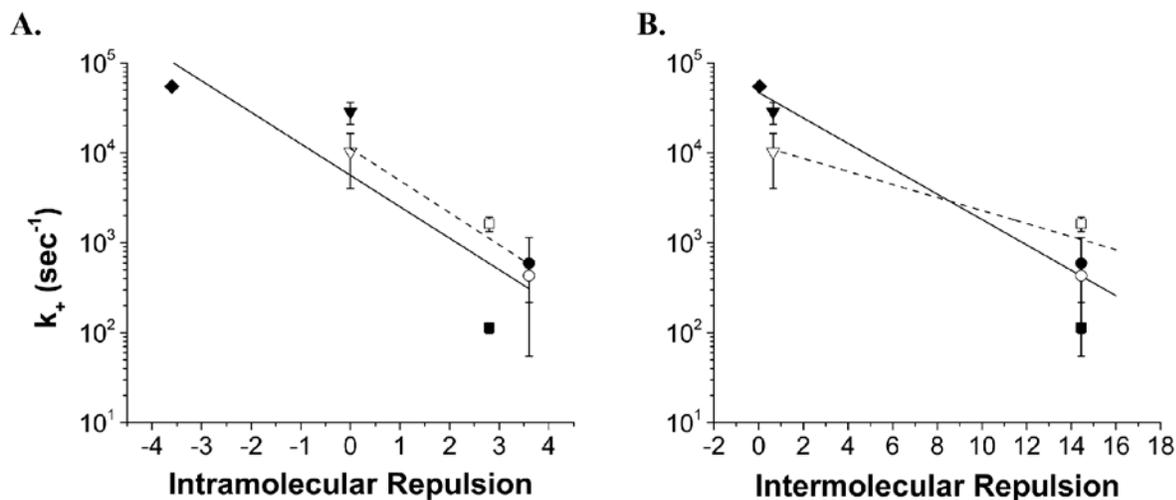
Peptide	$C_r, \mu\text{M}$	$\Delta G_{\text{elong}}$ (kJ/mol)	$k_+$ ( $\text{M}^{-1}\text{sec}^{-1}$ )	$n^*$	$\log(1/2 k_+^2 K_{n^*})$	$K_{n^*}$ ( $\times 10^{-9}$ )
$\text{K}_2\text{Q}_{26}\text{K}_2$	0.96	-33.98	$1.63 \times 10^3$	1.07	-3.15	0.56
$\text{KQ}_{26}\text{K}_3$	1.31	-33.20	$4.32 \times 10^2$	0.69	-3.05	14
$\text{Q}_{26}\text{K}$	0.66	-34.87	$1.03 \times 10^4$	1.01	-1.39	1.4
$\text{KQ}_{26}\text{D}$	0.77	-34.52	*	1.96	3.16	*
$\text{K}_2\text{Q}_9\text{P}^{\text{D}}\text{GQ}_9\text{K}_2$	2.30	-31.84	$1.13 \times 10^2$	0.81	-4.35	7.0
$\text{KQ}_9\text{P}^{\text{D}}\text{GQ}_9\text{K}_3$	2.37	-31.77	$5.95 \times 10^2$	0.82	-4.32	1.1
$\text{KQ}_9\text{P}^{\text{D}}\text{GQ}_9\text{K}$	1.11	-33.63	*	1.82	-2.63	*
$\text{Q}_9\text{P}^{\text{D}}\text{GQ}_9\text{K}$	1.29	-33.25	$2.86 \times 10^4$	1.02	-0.61	0.67
$\text{KQ}_9\text{P}^{\text{D}}\text{GQ}_9\text{D}$	0.51	-35.51	$5.49 \times 10^4$	0.68	0.04	0.74
$\text{K}_2\text{Q}_9\text{P}^{\text{D}}\text{GQ}_9\text{D}_2$	0.43	-35.95	*	1.57	-0.90	*

\* Since these peptides did not demonstrate  $n^*=1$ , further analysis of their thermodynamic properties were not conducted, and these values were not acquired or calculated.

Correlation of  $\Delta G_{\text{elong}}$  with intermolecular repulsion (Figure 4-4 B) is indicative of the contributions flanking charges make to the ability of peptide molecules to come together in order to elongate a fibril. Both D-Pro-Gly and simple  $\text{Q}_{26}$  peptides show similar trends in stability when intermolecular repulsion is greater than zero. There is a small, but gradual stabilization of fibril structure as repulsion decreases, with both peptide series demonstrating changes in free energy of less than 2 kJ/mol. In the D-Pro-Gly series, for the two peptides that have no intermolecular repulsion (◆ and◆, Figure 4-4), there is a greater stability increase than would be expected given the correlation observed for peptides with repulsion. The complexity of this relationship between fibril stability and intermolecular repulsion may be indicative of the nature of fibril elongation—elongating monomers must dock and arrange themselves to fit a particular existing three dimensional lattice. The structural details of this arrangement are still uncertain, making it difficult to explain any resulting complexities in energetics.

### 4.2.3 Fibril Elongation

Fibril elongation rates can be used as indicators of how a given set of charges affects the ability of disordered monomer to encounter and bind to fibril growing ends (docking), and then assume the correct structure to allow addition of subsequent monomers (locking) [35, 36]. The fibril elongation rate ( $k_+$ ) is a second order elongation rate constant and is defined by a relation between two values,  $k^*$  and [growing ends] (1.3.5.1, 2.2.4, 2.2.5). **Table 4-2** contains the elongation rates determined for each of the peptides with typical fibril morphology. The elongation rates determined for these peptides all lie within the typical range for polyQ peptides ( $10^2$ - $10^4$   $M^{-1}sec^{-1}$ ), and well below the diffusion limited value for protein-protein interactions [18, 19, 85, 138].



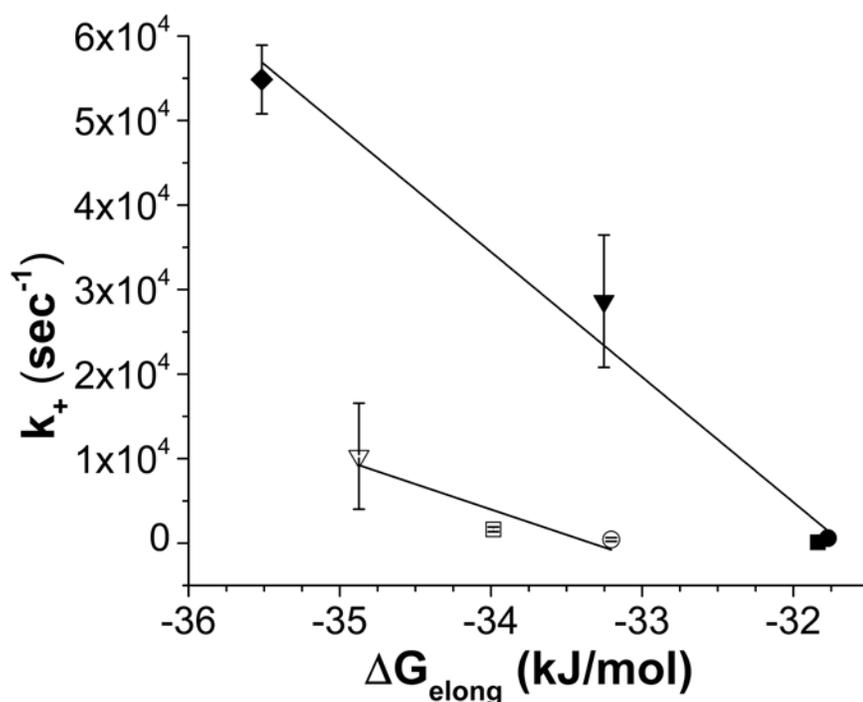
**Figure 4-5. Elongation rate constant vs. repulsion.**

Elongation kinetics correlations. Illustrates the relationship between the second order elongation rate constant and intramolecular (A) and intermolecular (B) repulsion.

As with  $\Delta G_{\text{elong}}$ , we found that  $k_+$  correlates well with intramolecular electrostatic repulsion: elongation rates in both peptide series increase as repulsion between the termini

decreases (**Figure 4-5 A**). This indicates that a decrease in repulsion between termini greatly enhances a peptide monomer's ability to encounter and elongate a fibril growing end. Similarly, there is a rate constant increase as intermolecular repulsion decreases (**Figure 4-5 B**). The corresponding trends in both inter- and intramolecular repulsions suggest that both intermolecular association and folding play important roles in fibril elongation. Enhancement of  $k_+$  could be due to a number of mechanistic factors: there could be improvements in the rate at which monomers encounter fibril ends, the rate at which they stably bind to these ends (dock), the rate at which they fold on the fibril surface to reconstitute a new fibril elongation face (lock), or any combination of the above. The experiments performed in this work did not allow us to determine which combination of steps might be most impacted by terminal charges.

When the fibril elongation rate constant is compared to fibril stability, these two values trend together—as fibril stability increases, the ability of that fibril to elongate also improves (**Figure 4-6**). This correlation is not surprising:  $\Delta G_{\text{elong}}$  is directly related to elongation rate through  $K_a$  ( $C_r^{-1}$ ), which is equal to  $k_+/k_-$ . When we consider that these two values were determined through entirely different methods, however, the good correlation is indicative of the quality of the data acquired by both approaches.



**Figure 4-6. Elongation rate constant vs. fibril stability.**

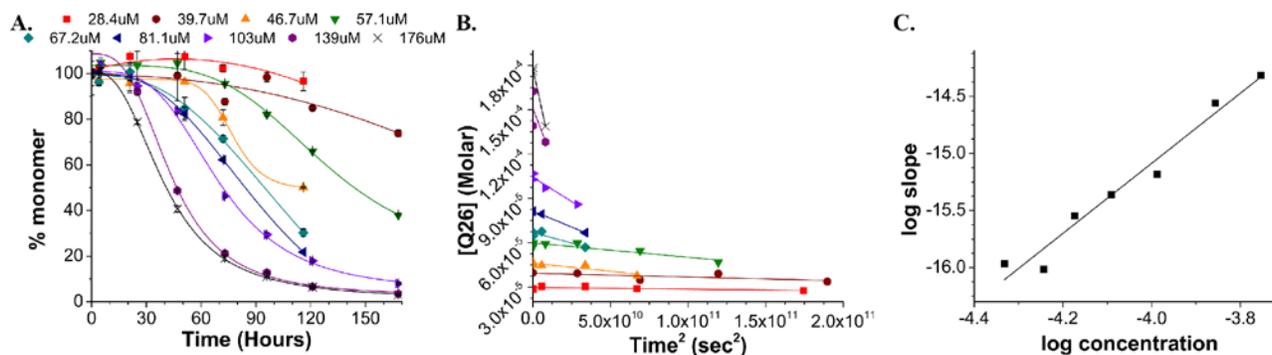
Fibril stability, as indicated by  $\Delta G_{\text{elong}}$  is directly related to elongation rate constant. The more stable a fibril (more negative  $\Delta G_{\text{elong}}$ ), the faster it is capable of elongating.

## 4.2.4 Application of Nucleated Polymerization Model

### 4.2.4.1 Nucleus size determination

While flanking charges have a demonstrable impact on polyQ fibril stability and elongation, their role in a peptide's ability to initiate fibril formation is still unclear. In an effort to gain some insight into this portion of the aggregation mechanism, we have conducted a nucleation analysis of all 10 peptides. To this end, the aggregation kinetics of each peptide were measured over a range of concentrations, as illustrated for  $K_2Q_{26}K_2$  in **Figure 4-7** (the concentration dependent kinetics for the other nine peptides are included in Appendix A). We were then able to apply

equations from the Ferrone thermodynamic model of nucleation (2.2.6) to derive essential energetic terms describing the nucleation and elongation mechanism of these peptides. The early time points of aggregation reactions were graphed versus  $\text{time}^2$  to yield a straight line, the rate of which corresponds to aggregation kinetics as described in Methods. The logs of these  $\text{time}^2$  rates are plotted versus the logs of the initial reaction concentrations to generate a log-log plot (**Figure 4-7 C**). The slope of the best fit line to this plot corresponds to  $n^*$  (equal to  $n^*+2$ ), while the y-intercept yields a complex parameter ( $\frac{1}{2}k_+^2K_{n^*}$ ) that relates the second order elongation rate constants for the nucleus and growing fibril to the nucleation equilibrium constant.



**Figure 4-7. Concentration dependent kinetics,  $\text{time}^2$  and log-log plots of  $\text{K}_2\text{Q}_{26}\text{K}_2$ .**

Time dependent aggregation kinetics for  $\text{K}_2\text{Q}_{26}\text{K}_2$  (A). From these kinetics, we are able to derive time squared plots of the first 20% of each aggregation reaction (B); these initial kinetics are comprised of a combination of fibril nucleation and the initial elongation phase. When time squared plots are fit with a best fit line, the rates can be used to generate a log-log plot, where the log slope of the time squared fit is plotted vs. the log of the reaction start concentration (C).

Most of the resulting  $n^*$  values indicate these peptides aggregate *via* monomeric nuclei, as expected given our intention when designing these peptides (1.4.1.1). Three peptides, however, result in an unexpected  $n^*\approx 2$ . These peptides,  $\text{KQ}_9\text{P}^{\text{D}}\text{GQ}_9\text{K}$ ,  $\text{K}_2\text{Q}_9\text{P}^{\text{D}}\text{GQ}_9\text{D}_2$ , and  $\text{KQ}_{26}\text{D}$ , also all had atypical morphologies present in the EMs (**Figure 4-2**). The presence of oligomers, protofibrils, or amorphous aggregates alongside fibrillar aggregates in the EM

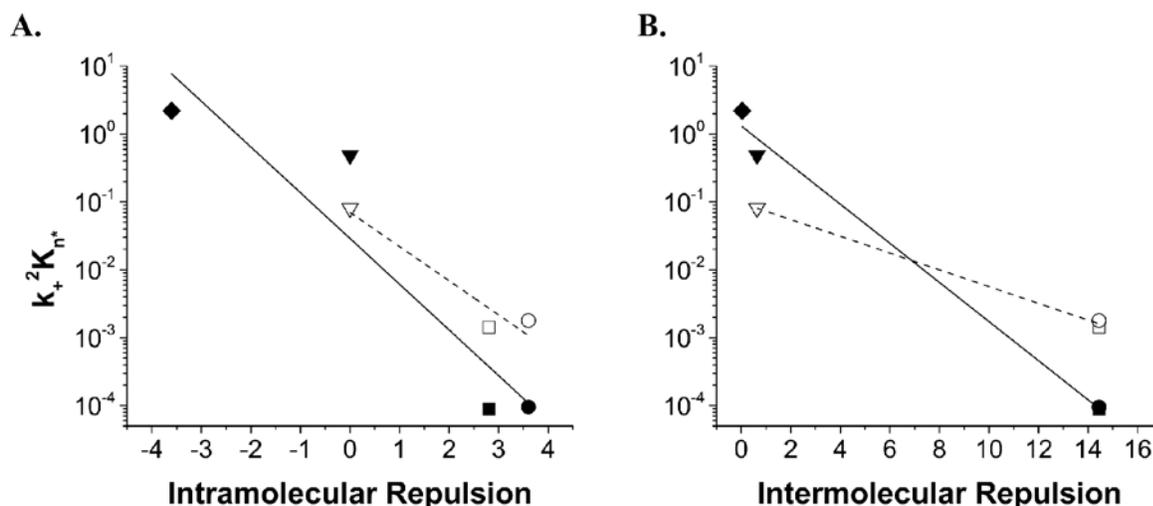
images, combined with the increase in nucleus size in the nucleation analysis together suggest a mixture of aggregation mechanisms. All of these peptides with  $n^* \approx 2$  share a common feature of low net charge. Two of these peptides have a net charge of -0.2, and the third a net charge of 1.8. It is likely that the decrease in solubilizing charge for these peptides plays a role in their ability to assume an alternative aggregation mechanism and form non-amyloidogenic aggregates. Worth noting is that some other peptides with low net charge (KQ<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>D and Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K), demonstrate no unusual morphologies or significant changes in  $n^*$ . So, while net charge does impact a peptide's ability to adopt these alternative mechanisms, low net charge is not a guarantee that multiple mechanisms will be present. Unfortunately, we were not capable of clarifying why these kinds of subtleties within net charge effects exist in this study.

Further support of the hypothesis that multiple mechanisms are in effect for these peptides comes from a comparison of high and low concentration reactions. For K<sub>2</sub>Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>D<sub>2</sub>, EMs of low concentration reaction products no longer showed the atypical aggregate morphologies seen in the high concentration reactions, but rather the tape-like morphology typically observed for polyQ aggregates. EMs of KQ<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K also demonstrate a morphology shift as concentration changes from high to low, though uncharacteristic aggregates are still present. These combined EM and mechanistic data suggest that the increased  $n^*$  values seen for some peptides reflect the operation of parallel mechanisms, where one (preferred at lower concentration) involves typical nucleated growth with  $n^*=1$ , and the other (favored at high concentrations) involves a higher order nucleation mechanism leading to the atypical morphologies observed by EM. While this hypothesis suggests that more uniform behavior might be obtainable for these low charge peptides by studies at lower concentrations, such low concentrations are difficult to study using our existing methods, and so were not considered here.

The remaining peptides in this study (with normal morphologies *via* EM) exhibit  $n^*$  values between 0.6 and 1 (**Table 4-2**), a typical range for peptides that aggregate *via* monomeric nuclei [18].

#### 4.2.4.2 Efficiency of spontaneous aggregation

The log-log y-intercepts of these 10 peptides, which are equal to  $\log(\frac{1}{2} k_+^2 K_{n^*})$ , range from 3.16 to -4.35, with more positive values reflecting more favorable aggregation properties (**Table 4-2**). The complex parameter derived from the y-intercept values of the log-log plot (equal to  $k_+^2 K_{n^*}$ ) contains the second order fibril elongation rate and the nucleation equilibrium constant. This complex thermodynamic parameter, along with  $n^*$ , captures the ability of a peptide to undergo spontaneous amyloid growth. Our analysis shows that the y-intercept trends with charge characteristics in much the same way as elongation rate and fibril stability. The correlations between this complex parameter and both inter- and intramolecular repulsion demonstrate that



**Figure 4-8. Complex parameter vs. repulsion.**

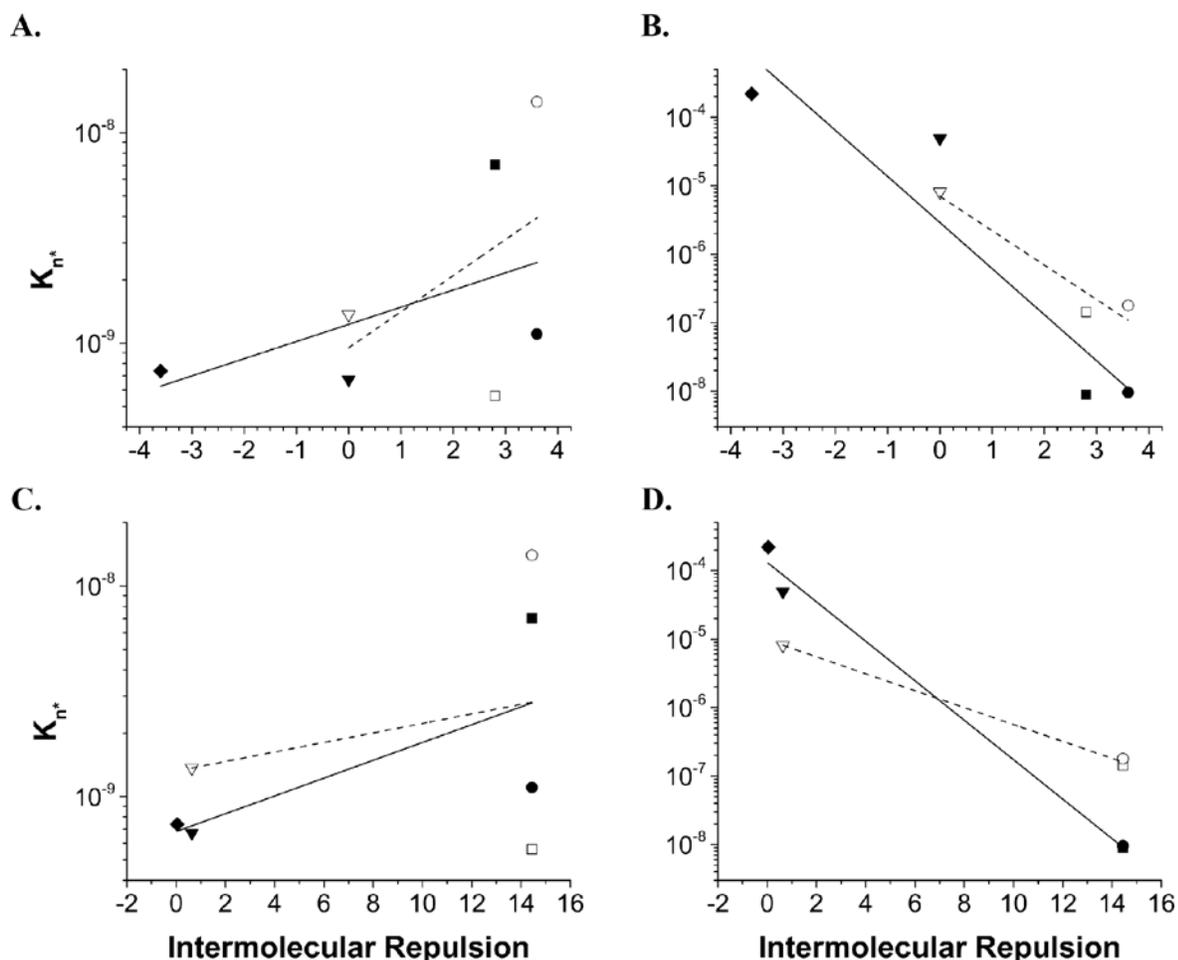
The trend of the complex parameter linking the elongation rate and  $K_{n^*}$  versus the intramolecular (A) and intermolecular (B) repulsion of the peptides.

the efficiency of nucleation among a series of peptides that all exhibit  $n^*=1$ , is enhanced by a decrease in terminal repulsion (**Figure 4-8**). This enhancement of fibrillization ability corresponds well with the previously described relationships between repulsion and aggregation rate or fibril stability.

#### 4.2.4.3 Nucleus stability calculations yield confusing results

By using  $k_+$  values in conjunction with the y-intercept from the log-log plot, it should be possible to calculate the  $K_{n^*}$  for each peptide (**Table 4-2, Figure 4-9 A and B**). These  $K_{n^*}$  values are extremely similar within each peptide series. Contrary to the known role of charges aiding in  $\beta$ -hairpin formation, our analysis suggests that the peptides with oppositely charged termini have slightly less stable nuclei than do those with repulsion between the termini.

This counter-intuitive result could be an artifact of the  $k_+$  value used to calculate  $K_{n^*}$ . This rate constant is determined using mature fibril seeds, and the mathematical approach used to calculate  $K_{n^*}$  assumes that the elongation rate of mature fibrils is identical to that of fibril nuclei and early phase aggregates (**Figure 1-6**). It is possible, however, that terminal charges would have a greater impact on elongation of a mature fibril than they would on a fibril nucleus, due to the substantial net charge of the fibril and the ease (or difficulty) by which an elongating monomer is able to approach the growing fibril. This difference in fibril charge perceived by the incoming monomer might prevent  $k_+$  for fibrils from matching the elongation rate of the nucleus.



**Figure 4-9.  $K_{n^*}$  vs. repulsion.**

The nucleation equilibrium constant,  $K_{n^*}$ , calculated from experimental (A,B) and constant (C, D)  $k_+$  values. The  $K_{n^*}$  values based on experimental measurements show only modest variation (only one order of magnitude between high and low values), and no clear trend indicating intra or intermolecular repulsion have an effect on nucleus stability. When one fixed elongation rate is used to calculate all  $K_{n^*}$  values, a clear trend emerges: the peptides with the least amount of repulsion form the most stable nuclei.

Continuing this train of thought, what happens if we assume that changes in terminal flanking sequences and net charge have little to no effect on either of the elongation rates, and therefore use a constant  $k_+$  for all peptides to calculate  $K_{n^*}$  (**Figure 4-9C and D**)? When calculated in this manner, nucleus stability trends as expected, with the hairpin-favoring attractive charges forming the most stable nuclei. Though this assumption leads to a pleasing

result, it cannot be independently justified. In fact, the trends in **Figure 4-6** suggest this assumption is not warranted. A more rigorous approach might be to allow  $k_+$  for fibrils to vary as measured, but to adopt an arbitrarily assigned, constant  $k_+$  for nuclei. Inspection of **Figure 4-9** suggests that, at best, this would leave  $K_n^*$  with some intermediate dependence on charge.

Because the best approach to calculate  $K_n^*$  is questionable, we have chosen to keep our focus on the complex parameter of  $k_+^2 K_n^*$  (**Figure 4-8**), which, along with  $n^*$ , dictates a peptide's ability to spontaneously form amyloid fibrils. The conclusions reached using this parameter are notable, and indicate that net charge and repulsion caused by charged flanking sequences play integral roles in polyQ aggregation.

### 4.3 DISCUSSION

These studies have allowed us to develop considerable insight into the qualitative and quantitative roles charged flanking sequences play in polyQ peptide aggregation. We hoped to answer a series of important questions about polyQ peptide solubility, and paint a picture of the role flanking charges might play in fibril structure, stability and aggregation mechanism. The results discussed above indicate that charges on the termini have definite impacts on the aggregation of polyQ.

Variations in terminal charge have an unmistakable qualitative effect on aggregation mechanism that appears to be related to net charge and its ability to solubilize polyQ. Low net charge tends to favor the development of a competing aggregation mechanism featuring an increase in nucleus size and leading to the formation of non-fibrillar aggregates. A relationship between the onset of this alternative mechanism and the high concentrations necessary for *in*

*vitro* experiments was suggested by both the increase in  $n^*$ , and the suppression of these mechanisms through decreases in reaction concentration. Whether the alternative mechanisms are brought about solely as a result of low net charge, and whether or not uncharged hydrophilic residues would discourage these mechanisms as effectually as charged residues, are still unclear. Reactions conducted in varying reaction conditions that result in a neutralization of net charge (**Appendix A.2**) further support the idea that neutralization of net charge can impact the aggregation of these peptides. However, this modification to aggregation behavior and our resulting inability to predict the exact performance of a peptide with charged flanking residues justifies our decision to be selective in the choice of flanking sequences and core peptides. The judicious selection of core polyQ peptides allowed for the generation of data that could be more easily interpreted and has led to a better understanding of how short amino acid sequences impact polyQ aggregation.

The quantitative impacts of terminal charged residues trended exactly as we might have expected considering their known ability to affect  $\beta$ -hairpin closure on other peptides, along with their roles in general intermolecular interactions. Energetic effects on polyQ fibril formation were relatively minor, with a maximal stability enhancement of approximately 10% when comparing the  $\Delta G_{\text{elong}}$  values. Elongation rate also trends well with decreases in repulsion. Variation in elongation rate is greatest for the D-Pro-Gly,  $\beta$ -hairpin mimetic series of peptides, which demonstrate increases in  $k_+$  of 2-3 orders of magnitude as repulsion decreases. Simple polyQ peptide elongation rates vary less, with the greatest  $k_+$  value only an order of magnitude higher than the lowest. The lesser impact on elongation rate for the simple polyQ peptides is likely more representative of the role flanking charges may have on disease polyQ proteins. Since disease polyQ proteins are longer and lack  $\beta$ -hairpin enhancing motifs, we might expect

further minimization in flanking sequence effects than those observed in the simple polyQ peptide. In any case, flanking sequences in general in disease proteins do not feature large collections of charged residues immediately neighboring the polyQ sequence.

Flanking charges are clearly important to aggregation mechanism and elongation, but what about nucleation? Unfortunately, any quantitative trends are lost when we calculate  $K_{n^*}$ : the resulting values have no discernible trend with net charge or repulsion. These calculated values may be accurate representations of the nucleation efficiency of these peptides. However, it is more likely that the standard assumption that elongation rates of nuclei are comparable to those measured for mature fibrils is flawed, as least when used to dissect charge effects, and contributed to these unexpected results. While this assumption may have little impact when comparing  $K_{n^*}$  values across a single flanking charge sequence (such as the  $K_2$ - $K_2$  flanking residues traditionally used in polyQ studies), variation of net charge and repulsion due to modification of flanking charges could result in substantial changes in  $k_+$  as compared to  $k_{+n^*}$  (**Figure 1-6**). The lack of a clearly defined relationship between elongation rates as terminal charge varies could explain our unexpected results for  $K_{n^*}$ , and suggests that any significant conclusions should not be based these data without further investigation.

In spite of the challenges in deciphering the nucleation equilibrium constants, the complex parameter,  $k_+^2 K_{n^*}$ , is still interpretable, and confirms that, in cases where  $n^*$  is constant, spontaneous aggregation efficiency improves with decreasing repulsion. The changes in this complex parameter are greater for the  $\beta$ -hairpin encouraging series than the simple polyQ peptides, again validating the design of our model peptide series and confirming the important impact that chain flexibility has on modulating charge effects on elongation and nucleation.

While charged flanking sequences clearly have an impact on the aggregation of polyQ peptides, this study demonstrates that our approach of adding two Lys residues to each polyQ terminus (K<sub>2</sub>-K<sub>2</sub> peptides) is a valid choice for the study of simple polyQ aggregation. This flanking sequence resulted in aggregation via a homogeneous monomeric nucleated polymerization mechanism, and led to the formation of fibrils with typical morphologies (as observed by EM). The use of a single, consistent flanking sequence across peptide studies also means that it is possible to compare K<sub>n</sub>\* values for all of these peptides, an important comparison to make when considering the impact of variations in the polyQ sequence on the aggregation mechanism.

How do these conclusions now translate to our understanding of polyQ in disease? These results make it clear that even a difference as small as 4 to the net charge of a peptide is capable of having a meaningful effect on aggregation, with notable effects on the fibril stability and elongation characteristics of a given peptide. However, our ability to translate these conclusions to the impact longer flanking sequences have on polyQ aggregation in disease is still limited. One takeaway may be that regardless of the flanking sequence, it must contain enough solubilizing charge to allow the polyQ sequence to remain soluble until amyloid formation is initiated. The 17 and 29-amino acid N-terminal flanking sequences of synthetic huntingtin exon 1 and AT7 model peptides have net charges of +1.5. Both of these peptides aggregate cleanly to form homogeneous amyloid, indicating that the small net charge imparted by these N-terminal fragments is clearly sufficient to prevent precipitous, disordered aggregation. It is worth noting that polyQ disease proteins have flanking sequences with a variety of charge densities and amino acids types. In addition, it is not clear how to think about the “solubilizing effect” of flanking

sequence on polyQ, when the flanking sequences are long sequences encompassing many domains and many types of sequences.

While flanking charged residues clearly play an important role in the aggregation of polyQ peptides, their contributions do not encompass the entire role flanking sequences have in the aggregation of polyQ disease proteins. Thus, while polyQ with AT7 flanking sequences maintains the simple polyQ aggregation mechanism [18], the critical early state of the aggregation mechanism of polyQ with htt-exon 1 flanking sequences is dominated by initial  $\alpha$ -helix formation and interactions between the N-termini [96]. In spite of such complications, the similarity in aggregation properties across terminal charge mutations found here further justifies our application of simple polyQ aggregation studies to amyloid formation in disease, since there is a well-defined dominating mechanism. Thus, even with disease proteins where flanking sequence imposes more elaborate, multi-step nucleation mechanisms, it appears—for example judging by repeat length dependence—that the fundamental polyQ aggregation mechanism and its sequence constraints continue to contribute to the overall mechanism [139].

## 5.0 REPEAT LENGTH DEPENDENT EFFECTS ON POLYQ AGGREGATION

### 5.1 OVERVIEW

Though it is well known that Gln repeat length plays an important role in amyloid nucleation [18] and that aggregation rate increases with repeat length [64, 74], little is known about the details of how additional glutamines contribute to fibril stability and the kinetics of fibril formation. By applying a nucleated polymerization model of amyloid formation [22] to the aggregation kinetics of polyQ peptides with a range of repeat lengths, it was possible to determine the role repeat length trends play in polyQ amyloidogenesis.

A series of polyQ peptides were chosen over a range of repeat lengths, and their aggregation kinetics and fibril stability were investigated (**Table 5-1**). Most peptides examined here lie above the monomeric nucleation threshold of 24-26 Glns (1.3.4), but two peptides that aggregate *via* tetrameric nuclei were also considered in an effort to provide some insight into the reason we observe this repeat length dependent shift in nucleus structure.

**Table 5-1. Repeat length peptides and features.**

Peptide Sequence	n*	log (1/2 k <sub>+</sub> <sup>2</sup> K <sub>n*</sub> ) <sup>a</sup>		C <sub>r</sub> (Molar)	ΔG <sub>elong</sub> (kJ/mol)
K <sub>2</sub> Q <sub>18</sub> K <sub>2</sub>	3.7 <sup>b</sup>	2.71		3.65E-6 <sup>c</sup>	-32.27
K <sub>2</sub> Q <sub>23</sub> K <sub>2</sub>	3.9 <sup>b</sup>	6.77		3E-6 <sup>c</sup>	-32.78
K <sub>2</sub> Q <sub>26</sub> K <sub>2</sub>	1.04	0.9 <sup>b</sup>	-3.15    -3.33	9.6E-7	-35.71
K <sub>2</sub> Q <sub>27</sub> K <sub>2</sub>	1.22	0.9 <sup>b</sup>	-3.63    -3.08	8.4E-7	-36.19
K <sub>2</sub> Q <sub>29</sub> K <sub>2</sub>	0.79	-2.48			
K <sub>2</sub> Q <sub>30</sub> K <sub>2</sub>	0.66	-2.8			
K <sub>2</sub> Q <sub>31</sub> K <sub>2</sub>				8E-7	-36.19
K <sub>2</sub> Q <sub>34</sub> K <sub>2</sub>	0.96	-1.68			
K <sub>2</sub> Q <sub>37</sub> K <sub>2</sub>	0.90	-2.09			
K <sub>2</sub> Q <sub>42</sub> K <sub>2</sub>	0.53	-0.92		3.5E-7	-38.33
K <sub>2</sub> Q <sub>47</sub> K <sub>2</sub>	0.87 <sup>d</sup>	-0.22			

<sup>a</sup> Calculated for all peptides except Q<sub>18</sub> and Q<sub>23</sub> with n\*≠1. <sup>b</sup> From [18]. <sup>c</sup> Personal communication from KK. <sup>d</sup> From [85].

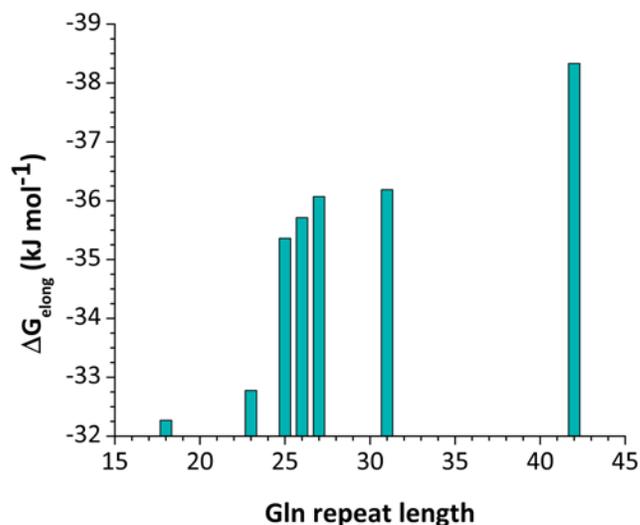
## 5.2 RESULTS

### 5.2.1 Fibril stability

C<sub>r</sub> was determined for each peptide as described previously (2.1.5). Comparisons of these values (Table 5-1) and their corresponding ΔG<sub>elong</sub> indicate a clear link between fibril stability and increasing repeat length (Figure 5-1). There is a notable enhancement of stability as the repeat length threshold between tetrameric and monomeric nuclei is crossed near Q<sub>25</sub>. Beyond this

threshold, the fibril continues to stabilize with increasing Gln length, but less and less with each subsequent Gln addition.

The differences in free energy of elongation across the transition from tetrameric nucleus to monomer, between Q<sub>23</sub> to Q<sub>26</sub>, are not insignificant.  $\Delta\Delta G_{\text{elong}}$  between these two peptides is 2.9 kJ/mol: an increase in stability of nearly 10% for an addition of only 3 residues, or nearly one kJ/mol per residue. While the  $\Delta\Delta G_{\text{elong}}$  between Q<sub>26</sub> (the shortest monomeric nucleating polyQ peptide) and Q<sub>42</sub> is comparable at 2.97 kJ/mol, it occurs over a span of 16 Glns, for an incremental increase of less than 0.2 kJ/mol per residue. These data trends suggest the possibility of a structural connection between the fibril stability trends and the switchover in nucleus size.

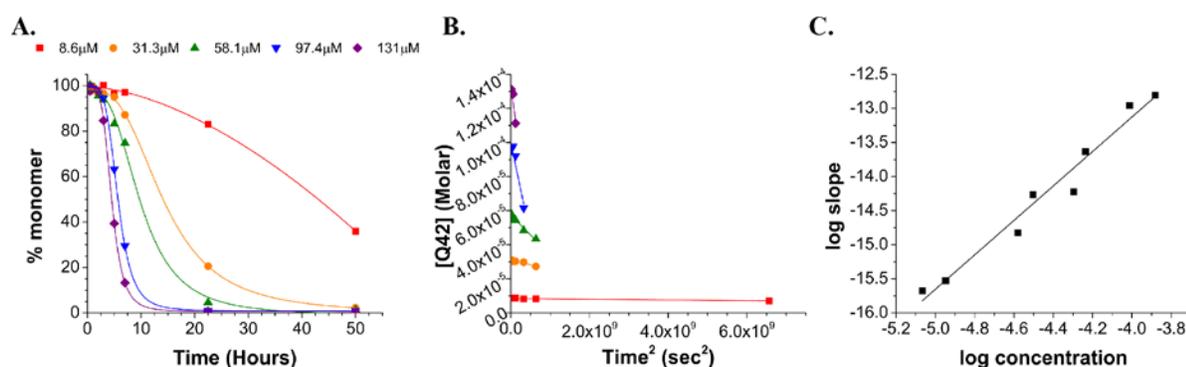


**Figure 5-1.  $\Delta G_{\text{elong}}$  versus repeat length.**

As Gln content of a polyQ peptide increases, so does fibril stability as indicated by  $\Delta G_{\text{elong}}$ . There is a notable jump in stability near Q<sub>25</sub>, suggesting aggregation through a monomeric nucleus results in a fibril 10% more stable than a fibril formed through tetrameric nucleus.

## 5.2.2 Spontaneous aggregation and nucleation

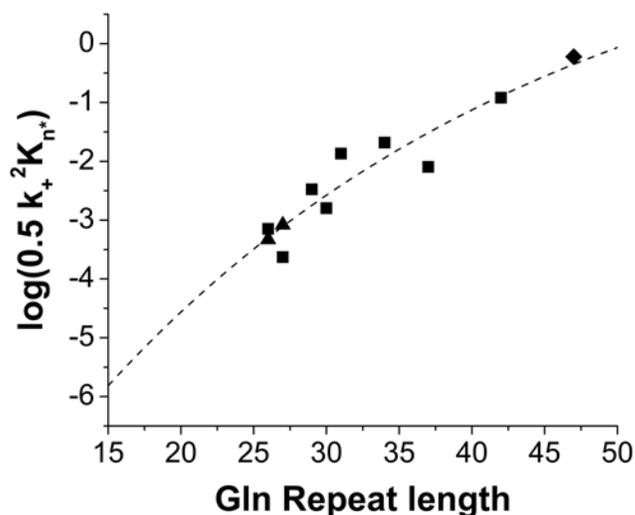
Aggregation mechanism and kinetics were investigated for all peptides by applying the Ferrone model of nucleated polymerization [22] (2.2.6). Aggregation kinetics were followed for a given peptide at a range of concentrations (shown for Q<sub>42</sub> in **Figure 5-2 A**), and the mathematical model was then applied to these data to generate time<sup>2</sup> (**Figure 5-2 B**) and log-log plots (**Figure 5-2 C**), as described (2.2.6) (data for other peptides is shown in Appendix section A.3). Nucleus size ( $n^*$ ) was determined from the best fit line of the log-log plot. The y-intercept, a complex parameter which is defined by  $k_+$  and  $K_{n^*}$ , was extrapolated based on a fit line with the slope fixed to 3, corresponding to  $n^*=1$  (**Table 5-1**). Fixing the slope to an integer value in this manner minimized the impact of experimental variability in fractional values of  $n^*$  on the derived y-intercepts. The best fit line of a log-log plot corresponding to a monomeric nucleus often yields a slope slightly less than 3 (fits slope of 2.6-3.1 can still correspond to monomeric nuclei), and this variation is entirely consistent with experimental discrepancies rather than an actual mechanistic effects that might lead to a fractional  $n^*$  [18].



**Figure 5-2.** Q<sub>42</sub> kinetics, time<sup>2</sup> and log-log plots.

In order to investigate the contributions of repeat length to nucleus stability and elongation rates ( $K_{n^*}$  and  $k_+$ ), the log-log y-intercept values, equal to  $\log(1/2k_+^2K_{n^*})$ , were plotted

versus Gln repeat length for peptides with  $n^*=1$  (**Figure 5-3**). This complex parameter corresponds to the ability of a given peptide with  $n^*=1$  to undergo spontaneous aggregation. These y-intercepts exhibit what appears to be a non-linear, possibly exponential trend with repeat length for polyQ peptides with monomeric nuclei. This trend is qualitatively similar to that of  $\Delta G_{\text{elong}}$  with repeat length. This suggests that peptides with the most stable end stage fibrils are also more favored to spontaneously nucleate and elongate to form these structures.

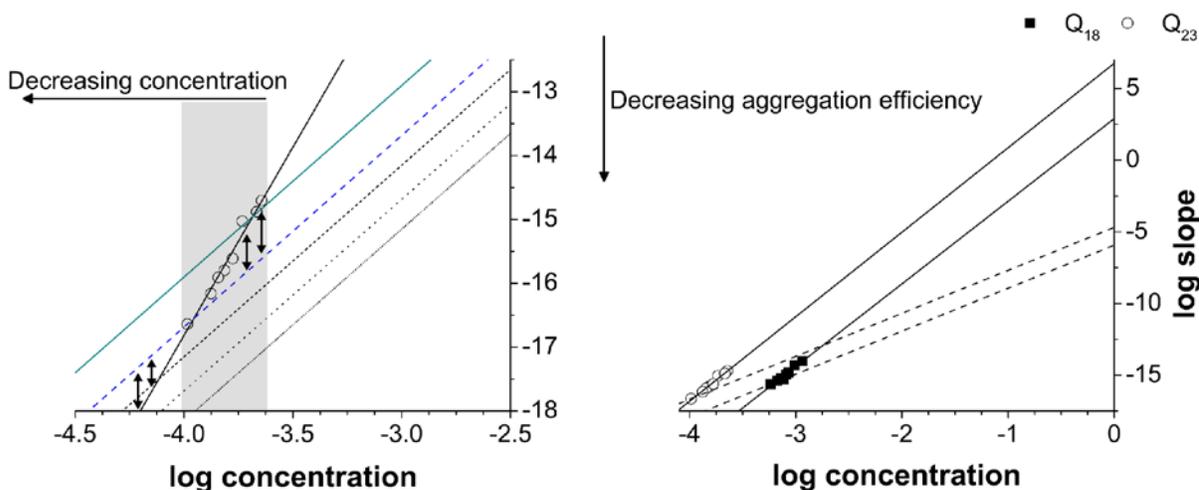


**Figure 5-3. A fit of y-intercept versus repeat length for peptides with monomeric nuclei**

The log-log y-intercepts, equal to  $\log(\frac{1}{2}k_+^2 K_{n^*})$ , of peptides with a monomeric nucleus ( $n^*=1$ ) as calculated from both new experiments (■) and previously published data (▲,◆), trend well with repeat length. The fit line is the best fit of the all data, both previously published and acquired for this study.

Unfortunately, this dataset and its fit leave much to be desired. If it is accurate at polyQ repeat lengths lower than  $Q_{26}$ , the fit indicates that  $Q_{23}$  should aggregate *via* a monomeric nucleus with only a modest drop-off in efficiency when compared to  $Q_{26}$  (as determined by the  $\log(\frac{1}{2}k_+^2 K_{n^*})$  value of the fit at these repeat lengths). As discussed, previous studies demonstrated that  $Q_{23}$  aggregates *via* a tetrameric nucleus [18]. However, the fit from **Figure 5-3**

suggests that we should have observed aggregation of  $Q_{23}$  via a monomeric nucleus in the reactions previously studied experimentally (located in the grey shaded region in **Figure 5-4 A**). This conclusion can be reached by examining the plots in **Figure 5-4**, which were generated by first plotting the actual experimental data (from [18]) for  $K_2Q_{18}K_2$  and  $K_2Q_{23}K_2$  in a log-log plot (experimental points plotted with their best fit line in solid black). In **Figure 5-4 A**, the log-log line for a monomeric nucleus (slope = 3) with the y-intercept predicted for  $K_2Q_{23}K_2$  by the fit in **Figure 5-3** (solid teal line) shows that the expected aggregation rates for this peptide at the concentrations tested in lab would have been faster than those observed experimentally.



**Figure 5-4. Calculation of monomeric y-intercepts for short polyQ peptides.**

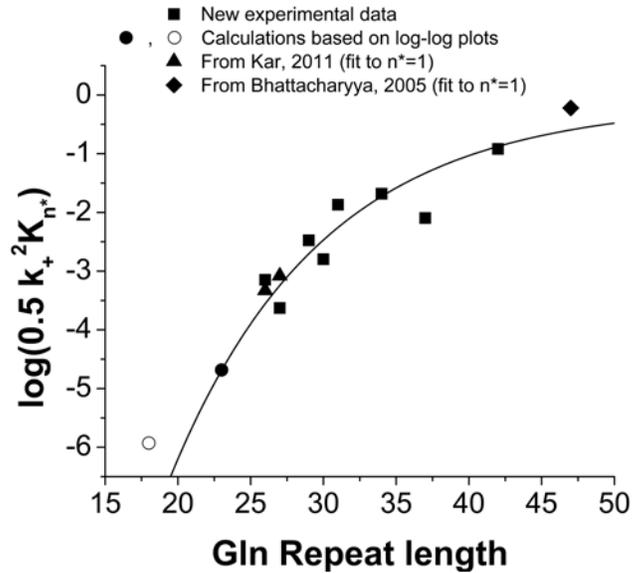
Experimental log-log plots of  $Q_{23}$  ( $\circ$ ) (A, B) and  $Q_{18}$  ( $\blacksquare$ ) (B) and their best fit lines (solid black). The theoretical log-log plot based on the fit in **Figure 5-3** is shown in teal in (A). When compared to the experimental data, this line indicates experiments should have measured a monomeric nucleus, rather than the tetramer, actually observed. When a hypothetical fit for  $n^*=1$ , slope=3, (dashed line) is fit through the lowest concentration point of each of these data, a maximal y-intercept can be extrapolated (B, blue dashed line in A). It is also possible that the spontaneous aggregation efficiency of these peptides through a monomeric nucleus is lower than that indicated by our maximal y-intercept. This would result in log-log fit lines lower than the one approximated here (for example, black dotted lines) (A).

The most likely explanation for why aggregation via a monomeric nucleus was not observed is that the y-intercept value for monomeric nucleation by  $Q_{23}$  is overestimated in the **Figure 5-3** fit. We do believe it is possible, under the right conditions, for some of these short peptides to achieve monomeric nuclei; they are simply less energetically favored than the tetrameric nucleus under the experimental conditions by which they have been studied (if there are enough residues to form a  $\beta$ -hairpin structure, aggregation *via* this structure as a monomeric nucleus should be possible). In an effort to gain a better understanding of the conditions necessary to achieve a monomeric nucleus for peptides with repeat lengths below the  $Q_{25}$  threshold, we generated theoretical monomeric log-log plots for  $Q_{23}$  and  $Q_{18}$  based on previous experimental data (**Figure 5-4**). The effects of various theoretical y-intercept values on the concentration dependence of aggregation via a monomeric nucleus were investigated by drawing lines with slopes of 3 (i.e., for  $n^*=1$ ) originating at various y-intercepts. By forcing one of these lines through the point corresponding to the lowest peptide concentration studied (the concentrations at which monomeric nuclei are favored must be lower than the lowest concentration point that fits the  $n^*\approx 4$  line, otherwise aggregation *via* a monomeric nucleus would occur at a faster rate than aggregation *via* a tetramer) we have generated log-log fits that produce maximal y-intercept values for a monomeric nucleus within each of these peptides (e.g. blue dashed line in **Figure 5-4 A**). The difference in the rates of the tetrameric nucleus and the monomeric nucleus with maximal y-intercept are illustrated by double sided arrows in **Figure 5-4 A**. It is possible (and even likely) that these low experimental concentrations are significantly higher than the actual concentration at which a monomeric nucleus could be favored, so (if a monomeric nucleus does exist for these peptides) the actual log-log fit would be a line of slope

three that is no higher than those plotted in **Figure 5-4 B** (as illustrated for Q<sub>23</sub> by the black dashed or dotted lines in **Figure 5-4 A**).

When these theoretical y-intercept upper limits are added to the plot of  $\log(\frac{1}{2}k_+^2K_n^*)$  versus Gln repeat length (● and ○ in **Figure 5-5**), and the fit is forced through the Q<sub>23</sub> point, the exponential trend from **Figure 5-3** becomes more defined. This new fit indicates removal of additional Gln residues results in notable decreases in the ability of a polyQ peptide to spontaneously aggregate through a monomeric nucleus. Thus, Q<sub>23</sub> and Q<sub>18</sub> are not observed to aggregate *via* monomeric nuclei because of the steep drop in nucleation efficiency in the range of Q<sub>25</sub>, and the existence of a competing tetrameric nucleus pathway.

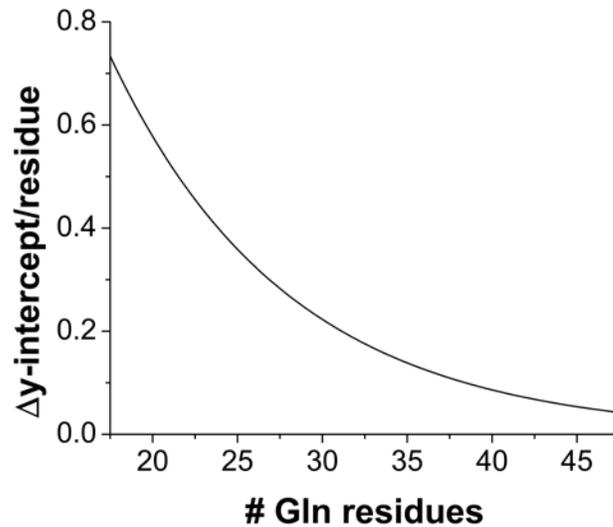
In fact, the change in y-intercept per residue ( $\Delta y\text{-int/res}$ ) increases rapidly at repeat lengths lower than 25 (**Figure 5-6**), and suggests the actual Q<sub>18</sub> y-intercept is likely to be significantly lower than the approximated maximum from **Figure 5-4**. According to this fit, Q<sub>18</sub> peptides favor a monomeric nucleus at concentrations less than 177  $\mu\text{M}$ —a reaction start concentration nearly 3 times lower than the lowest concentration previously studied [18]. Any reactions conducted at these concentrations would be extremely slow due to their inefficient aggregation ability. The lowest concentration followed thus far, 580  $\mu\text{M}$ , has a  $t_{1/2}$  of approximately 200 hours; decreasing the concentration by a factor of 3 would only further slow the kinetics, making it extremely difficult to observe any aggregation.



**Figure 5-5. New fit of y-intercept vs. repeat length for peptides with monomeric nuclei**

When an exponential fit for all shaded data points is forced through the projected  $Q_{23}$  y-intercept value (●), we generate a fit that appears to match the efficiencies of aggregation of various polyQ peptides as they aggregate through monomeric nuclei.

While this new fit is very convenient, and it explains nicely why we do not see monomeric nuclei for short polyQ lengths, it is purely conjectural. It would be ideal if we could somehow test whether this fit is a valid predictor of the y-intercept for short polyQ peptides. We have previously discussed the experiments conducted with two short D-Pro-Gly peptides,



**Figure 5-6.  $\Delta y$ -int/residue versus repeat length.**

When tangents from the fit line from Figure 5-5 are used to calculate  $\Delta y$ -int/res, we find the y-intercept changes more per residue for short polyQ peptides, indicating that removal of a single Gln in short peptides has a greater impact on aggregation ability than on long peptides.

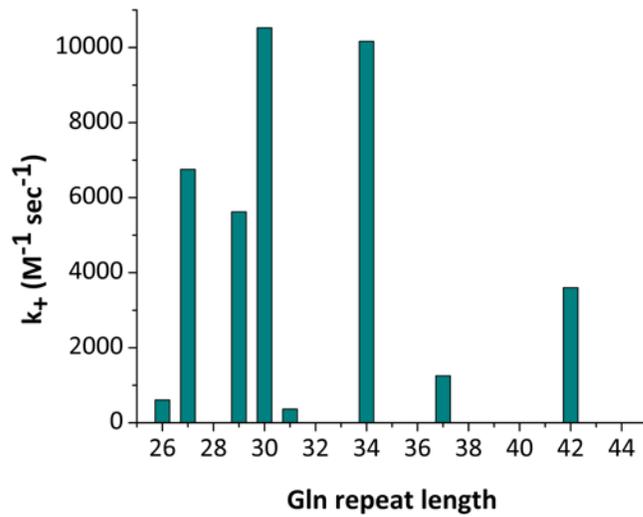
$K_2Q_9P^D GQ_9K_2$  and  $K_2Q_{10}P^D GQ_{10}K_2$ , which have  $n^*=1$  (1.4.1.1). Conveniently, the experimental log-log plots of these hairpin peptides can be used to generate a  $\Delta y$ -int/res values that can be used to test the reasonableness of the fit in **Figure 5-5**. The  $\Delta y$ -intercept between these  $Q_{20}$  and  $Q_{22}$  analog peptides is 0.98, which corresponds to a  $\Delta y$ -int/res of 0.49 for a  $Q_{21}$  peptide. As can be seen in **Figure 5-6**,  $\Delta y$ -int/res from the **Figure 5-5** fit varies along polyQ repeat length from  $\sim 0.1$  to  $\sim 0.7$  (for repeat lengths that have been studied experimentally). According to this figure, the theoretical  $\Delta y$ -int/res of  $Q_{21}$  is 0.53, an excellent match to the experimental value determined from the hairpin peptides. This consistency between  $\Delta y$ -int/res values indicates the fit line in **Figure 5-5** is actually a fairly realistic approximation of the y-intercept of polyQ peptides, validating our approach. The steepness of the  $\Delta y$ -int/res plot in this repeat length range helps

explain the very sharp transition in experimental  $n^*$  values and ultimate mechanism change at low repeat length.

### 5.2.3 Elongation rates and nucleation equilibrium constants

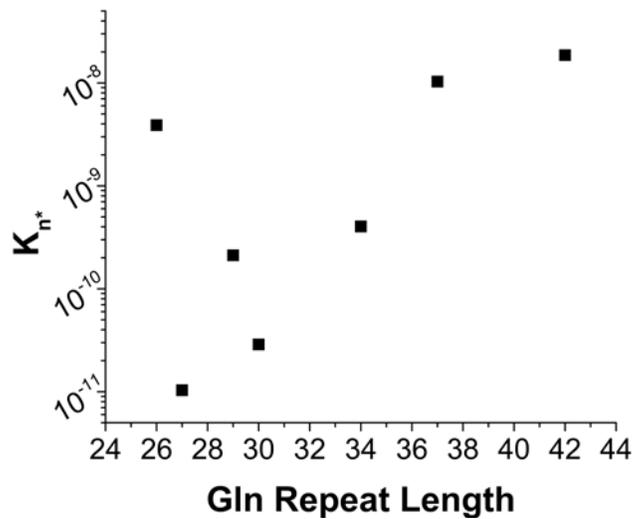
The elongation rate,  $k_+$ , for each peptide was determined using  $k^*$  and [growing ends] values from seeded elongation reactions and growing ends titration assays, as described (2.2.4, 2.2.5). The resulting elongation rates are plotted in **Figure 5-7**. These  $k_+$  values vary significantly more than expected given the similarities in previously reported polyQ  $k_+$  values [18, 85]. These new  $k_+$  values have no clear trend with polyQ length, and in fact appear to vary periodically as repeat length increases.

According to the Ferrone equation, the  $k_+$  values and the log-log y-intercepts can be used in conjunction to determine the  $K_{n^*}$  for these peptides (**Figure 5-8**). In spite of the large variation in  $k_+$ , the  $K_{n^*}$  values still demonstrate a rough correlation with polyQ repeat length. As the number of Gln residues increases,  $K_{n^*}$  also increases, demonstrating that, as might be expected, addition of Gln residues improves the stability of the monomeric nucleus structure.



**Figure 5-7.  $k_+$  versus repeat length.**

Elongation rate varies significantly over the polyQ peptides studied, and has no clear trend with repeat length.



**Figure 5-8.  $K_{n^*}$  versus repeat length.**

Nucleus stability as indicated by  $K_{n^*}$  increases with Gln content. Significant changes in  $K_{n^*}$  occur when repeat length is in the mid-thirties, consistent with disease threshold length.

### 5.3 DISCUSSION

We investigated a series of polyQ peptides across a range of repeat lengths (Q<sub>18</sub>-Q<sub>42</sub>) to determine the resulting trends in fibril stability, elongation ability, and nucleation. The data confirm that addition of Gln residues increases fibril and nucleus stability, as previously posited [74]. More interestingly, the number of Gln residues in a polyQ sequence has a predictable effect on these terms, which allows for the estimation of kinetics parameters and thermodynamic values for polyQ repeat lengths that have not been measured.

All three terms with values that corresponded well with Gln repeat length ( $\Delta G_{\text{elong}}$ ,  $y$ -intercept, and  $K_n$ \*) have transition points, or repeat length thresholds that indicate subsequent increases in Gln content result in only minor variations in their value. For  $\Delta G_{\text{elong}}$ , this threshold corresponds well with the threshold for nucleus size, in the mid-twenties. This suggests that there exists some factor that is common to both formation of a stable monomeric nucleus and fibril stability, possibly the ability to favorably fold into a  $\beta$ -hairpin structure. The log-log  $y$ -intercepts described in this chapter indicate the efficiency with which a given polyQ can spontaneously aggregate through a monomeric nucleus. This efficiency appears to increase less markedly once repeat length enters the 30's (**Figure 5-6**). This correspondence to pathogenic repeat length indicates that enhancement of a polyQ sequence's aggregation ability beyond its capacity to form a monomeric nucleus may play a key role in the threshold of disease.

For the first time, it has been possible to confirm the feasibility of monomeric nuclei in peptides that experimentally exhibit multimeric nuclei, such as Q<sub>18</sub> and Q<sub>23</sub>. For these peptides, monomeric nuclei are demonstrated to be favored in low concentration reactions that aggregate extremely slowly, because they are so inefficient at aggregation (as indicated by their low  $y$ -intercepts): Q<sub>23</sub> peptides favor a monomeric nucleus at concentrations of 100  $\mu\text{M}$  or less, while

Q<sub>18</sub> aggregates *via* a monomeric nucleus when reaction concentration is less than 177 μM. By decreasing the likelihood for multiple molecules to associate (through decreases in concentration), aggregation through a tetrameric nucleus becomes a less favored mechanism. The point at which this switch in mechanism occurs is representative of a concentration dependent energetic transition.

In contrast with clear trends in  $\Delta G_{\text{elong}}$ ,  $K_{\text{n}^*}$  and the log-log y-intercept,  $k_+$  does not follow a pattern with increasing Gln repeat length. In fact,  $k_+$  even appears to oscillate as repeat length increases from Q<sub>26</sub> to Q<sub>42</sub>. This variation in the elongation rate constant may indicate subtle, and previously unobserved, repeat length effects on the manner in which polyQ peptides are able to complete the dock-and-lock mechanism of elongation. For example, the width of the  $\beta$ -strand may play an important role in elongation—if a peptide is too long or short to easily form  $\beta$ -strands of the ideal length, it may take longer to undergo the “locking” step of aggregation and have a decrease  $k_+$  compared to other peptides of similar length (such as Q<sub>37</sub> in **Figure 5-7**)

Continued improvements in nucleus and fibril stability beyond the disease threshold are consistent with the decrease in age of onset and more rapid deterioration of patients with longer polyQ repeat lengths, a trend that continues well above the repeat length threshold [41] (1.2.1). The results presented in this chapter suggest that the ability of a polyQ expansion to stably form and aggregate efficiently through a monomeric nucleus plays an important role not only in fibril stability, but also in disease initiation and progression.

## 6.0 CONCLUSIONS

### 6.1 THESIS CONTRIBUTIONS

When this work was first begun, the lab had just confirmed a  $\beta$ -hairpin structure as a likely candidate for the monomeric nucleus observed in aggregation of polyQ peptides of 26 Glns or more. As the polyQ aggregation mechanism was developed, questions arose regarding the contributions of portions of the primary polypeptide sequence to the energetics of this reaction [122] and the validity of our proposed mechanism indicating polyQ aggregates *via* a homogeneous nucleated growth polymerization mechanism with a monomeric nucleus and monomer addition [120]. The intent of this work was to gain more specific information regarding the roles of terminal charge and additional Gln residues in the aggregation mechanism of simple polyQ peptides.

#### 6.1.1 Use of growing ends titration assay

The experiments conducted in Chapter 3.0 indicate that the growing ends titration assay most likely functions through the detection of one added label of many incorporated onto the elongation site of a growing fibril. More specifically, we believe the terminal biotinyl-polyQ incorporated into a fibril maintains a biotin accessible to solution and labeled streptavidin, while internal biotins are somehow sequestered and not detectable.

When we used the results from this assay in Chapter 4.0 to determine  $K_{n^*}$  (4.2.4.3), the resulting values were surprising and counterintuitive. This may speak to limitations of the approach set forth in Ferrone's original mathematical model, in which he equates the elongation rate of the nucleus to that of the fibril (discussed in 1.3.5.1). While the assumptions used to set this equality may apply in cases where  $n^*$  is equal to some large multimer, they may not hold true when the electrostatic 'appearance' of a fibril to an incoming monomer is much different than that of the monomeric nucleus (which is likely the case for many of the peptides studied in Chapter 4.0 ). Therefore, while the growing ends titration may lead to a reasonable value for fibril elongation rate ( $k_+$ ) we cannot always use this value coupled with Ferrone's equations to determine nucleus stability. On the other hand, it seems likely that any discrepancies caused by the use of the  $k_+$  value for both equilibrium constants are minimized when comparing peptides with nearly identical electrostatics (i.e.  $K_2Q_{27}K_2$  versus  $K_2Q_9P^DGGQ_9K_2$ ), because their  $K_{n^*}$  values, if biased at all, will be biased by the same factor (the ratio between  $k_{+n^*}$  and  $k_+$ ). This effect does make it difficult to compare  $K_{n^*}$  across variably charged peptides, making our ability to interpret terms such as the complex parameter describing spontaneous aggregation ability (that is, the y-intercept of the log-log plot) all the more important.

### 6.1.2 Terminal charge contributions

Application of a model for homogeneous nucleated polymerization to the aggregation kinetics of polyQ peptides with varying terminal charges allowed us to parse out the contributions these charges have on the aggregation of polyQ peptides. The results suggest that solubility of the polyQ containing molecule is important to its ability to aggregate *via* a single nucleated polymerization mechanism—when solubility is decreased because of insufficient net charge, a

competing mechanism developed, resulting in the appearance of non-amyloidogenic morphologies *via* EM and an increase in nucleus size.

Among those polyQ sequences examined that exhibit homogeneous amyloid formation mechanisms and products, terminal charges also play an important role in a polyQ peptide's ability to spontaneously aggregate. The lower the amounts of inter- and intra-molecular repulsion (and/or the greater the amount of attraction), the more efficiently a peptide is able to spontaneously aggregate, and the more stable the resulting fibril. These trends suggest that the abilities to fold into a stable hairpin (bringing termini close together), as well as bimolecularly associate to elongate fibrils play similar and important roles in polyQ fibril formation.

This chapter also further validates our hypothesis that polyQ aggregates *via* a monomeric nucleus (for repeat lengths above a certain number) and elongates through monomer addition. Our ability to detect the presence of oligomeric and non-amyloid species in reactions of peptides with low net charge, both visually by EM and through application of our thermodynamic model, suggests that we have the general capability to identify them if and when they are involved in aggregation mechanisms. The lab's earlier work investigating the detailed mechanism of simple polyQ aggregation used multiple methods that together ruled out the presence of oligomers at the start and during the reaction (DLS in particular, but also 20 nm filtration tests and EM) [18]. Thus, we can say that under certain conditions (such as a decrease in net charge) simple polyQ peptides are capable of forming non-amyloid oligomers as end products of the aggregation reaction, however, they are not the bulk of the reaction product, and can usually be eliminated by decreasing reaction concentration. They also are not confirmed to be intermediates in the aggregation of polyQ amyloid, but may be the end stage product of aggregation *via* a separate mechanism, as indicated by the composite  $n^*$  values, and their presence in end stage reactions.

In general, the effects of charged flanking sequences trend well with what might be expected based on our proposed mechanism of aggregation for simple polyQ peptides. The role such charge effects might have when incorporated to longer flanking sequences (such as those in disease proteins) is still unclear, and requires future study. However, this work clearly indicates that flanking sequences can have unmistakable effects on polyQ aggregation.

### 6.1.3 PolyQ repeat length

The same approach was applied to a series of different repeat length polyQ peptides to further investigate the kinetic and thermodynamic components to polyQ repeat length thresholds observed in *in vitro* aggregation and *in vivo* pathogenesis. The resulting trends in fibril and nucleus stability, and in spontaneous aggregation ability, indicate that all of these terms experience enhancement with addition of Gln residues up to and through the pathological threshold range.

The mechanistic transition from a tetrameric to monomeric nucleus between  $Q_{23}$  and  $Q_{26}$  previously suggested a possible contributing factor to the biophysical basis for disease thresholds [18]. An extrapolation of aggregation ability (as indicated by the log-log y-intercept) of polyQ peptides with monomeric nuclei suggests that the transition between monomeric and multimeric nuclei exists at the intersection of two trend lines, each indicative of its own aggregation mechanism/nucleus size, that describe the concentration dependence of the energetics of the aggregation mechanism. Thus, there appears to be no intrinsic barrier to nucleation of amyloid formation via a monomeric nucleus for  $K_2Q_{23}K_2$  peptides in the experimentally accessible concentration range, but the mechanism is not observed because nucleation via a tetrameric nucleus is more favorable at these concentrations. A major contribution to this situation appears

to be a rather dramatic reduction in nucleation efficiency (as reported by y-intercepts) as repeat length decreases through the Q<sub>26</sub> to Q<sub>18</sub> range.

This large drop in nucleation efficiency is further support for a  $\beta$ -hairpin as the nucleus structure: as repeat length decreases in this range, each subsequent removal of a Gln that could contribute to hydrogen bonding leads to a larger and larger losses in stabilization of the hairpin nucleus, corresponding to larger and larger decreases in y-intercept.

Thus, even among polyQ sequences that share a unitary value for n\*, further lengthening of the polyQ repeat lengths leads to further improvements in a polyQ peptide's ability to spontaneously aggregate and stabilize amyloid fibrils. Both of these energetic consequences exhibit transition points at or near the pathological disease threshold, suggesting that the aggregation properties of polyQ peptides beyond this transition point are likely contributors to the disease state.

#### **6.1.4 Summary**

The results of these studies indicate that mutations in polyQ sequences that enhance their ability to both form  $\beta$ -hairpins and associate (whether by increases in repeat length or attraction) facilitate polyQ aggregation through a fibrillization pathway with a monomeric nucleus. In general, fibril stability corresponds well with spontaneous aggregation ability. These results could have implications for polyQ in disease—any mutation that enhances amyloid formation could result in a decrease in repeat length threshold or enhanced pathogenicity. For example, the SCA6 low repeat length threshold (**Table 1-1**) could be a result of enhancements to aggregation ability caused by  $\beta$ -hairpin stabilizing effects of the flanking sequences.

## 6.2 FUTURE DIRECTIONS

### 6.2.1 Does peptide net charge or solubility have a greater impact on mechanism?

Results described in 4.2.4.1 suggest that low net charge on simple polyQ peptides may introduce a preference for an alternative aggregation mechanism. Experiments could be conducted to confirm this hypothesis. One approach would be to use hydrophilic, but uncharged residues in an effort to provide solubility to the polyQ even in the absence of formal charges. The use of Ser or Thr, both relatively polar, uncharged residues, may aid enough in solubilizing the polyQ enough to allow monitoring of the aggregation kinetics. If these residues sufficiently solubilize polyQ sequences, concentration dependent kinetics could be measured, allowing for the confirmation (or contradiction) of the hypothesis that low net charge is a contributor to the assumption of alternative aggregation mechanisms.

### 6.2.2 Can we generate a monomeric nucleus from low repeat length peptides?

The data presented in 5.2.2 suggests that Q<sub>23</sub> should aggregate *via* a monomeric nucleus at low concentrations. If this is true, it should be possible to confirm by conducting reactions at the low concentrations hypothesized to generate an  $n^*=1$  for peptides shorter than Q<sub>25</sub>. These experiments have not been conducted up to this point because the low concentrations necessary for their success are time prohibitive. In addition, because of the very long incubation times required and relatively low aggregation extents expected, it is not clear that data of sufficient quantity can be obtained, even if one is willing to devote the necessary time and material to the experiment. As experimental methods continue to develop, our ability to accurately follow the

kinetics of these long time course reactions without the need for excessive material may improve. The successful generation of an experimental log-log plot for peptides of repeat lengths of  $Q_{23}$  or lower, similar the simulated plots described in Chapter 5.0 would confirm our conclusion that short repeat length polyQ peptides are capable of accessing a monomeric nucleus whose formation and/or elongation, however are energetically unfavorable.

### **6.2.3 Big picture**

While these studies have led to a clearer understanding of the kinetics and thermodynamics of aggregation in models of simple polyQ, it is still unclear how this mechanism translates to the aggregation of disease relevant peptides or proteins. Flanking sequences clearly have an impact on aggregation mechanism, whether by introduction of an alternative mechanism or enhancement of the existing one. To better evaluate whether differences in pathogenicity across polyQ disorders can be attributed to aggregation propensities, systematic studies of all disease relevant flanking sequences should be conducted. This work has already been begun by our lab with Htt<sup>NT</sup> and AT7 model peptides. By investigating other disease model peptides we might gain further insight into the changes in aggregation mechanism caused by longer flanking sequences, as well as the degree to which mechanistic details learned from simple polyQ peptides might be also relevant to disease protein aggregation.

## APPENDIX A

### SUPPLEMENTARY DATA

#### A.1 COMPLETE KINETICS DATA FOR TERMINAL CHARGE PEPTIDES

The kinetics,  $\text{time}^2$  and log-log plots for the nine terminal charge peptides not shown in Chapter 4.0 , Section 4.2.4.1. For all peptides, kinetic traces are shown in (A), the  $\text{time}^2$  plots for the early phase kinetics in (B), and the resulting log-log plots in (C). Only peptides with good fits ( $R^2 \geq 0.7$ ) to the  $\text{time}^2$  plot were used in log-log plots.

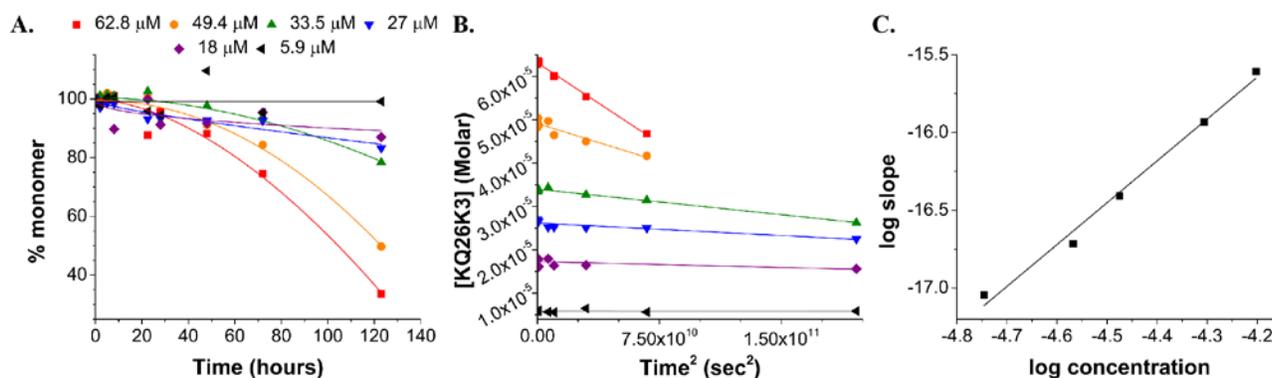


Figure A-1. KQ<sub>26</sub>K<sub>3</sub> kinetics,  $\text{time}^2$ , and log-log plot.

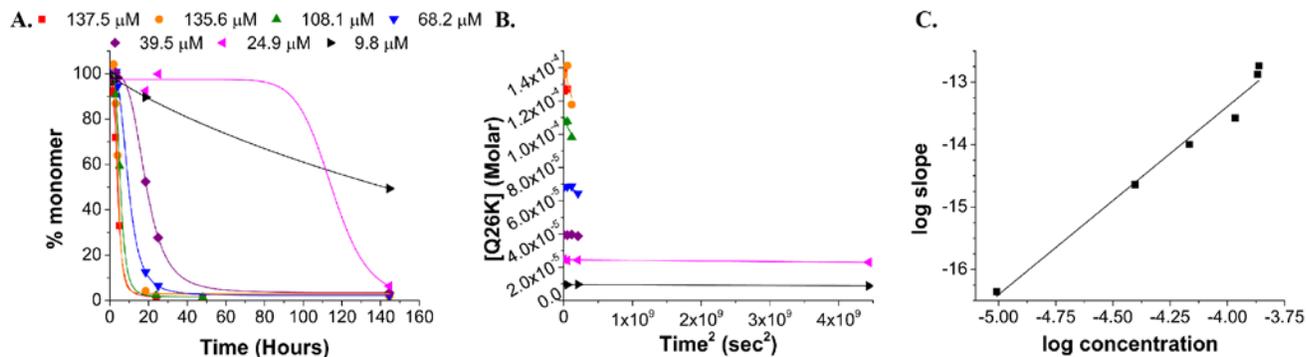


Figure A-2. Q<sub>26</sub>K kinetics, time<sup>2</sup>, and log-log plot.

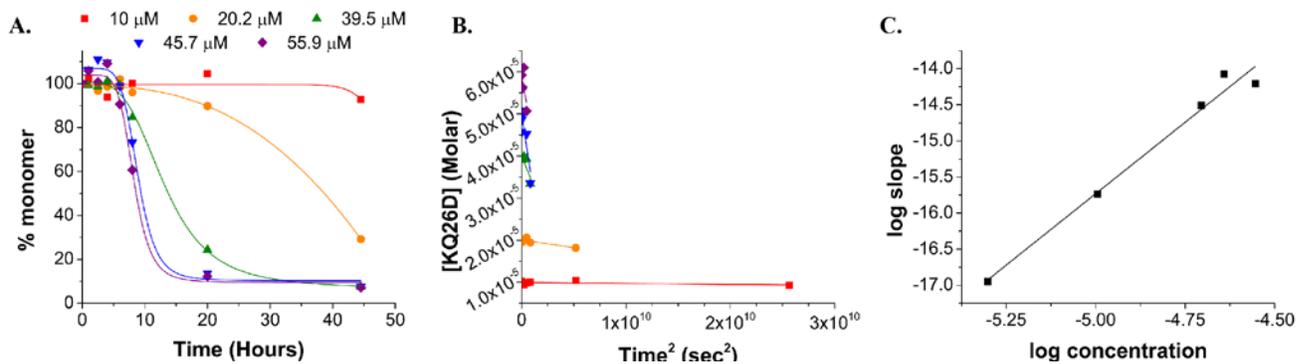


Figure A-3. KQ<sub>26</sub>D kinetics, time<sup>2</sup>, and log-log plot.

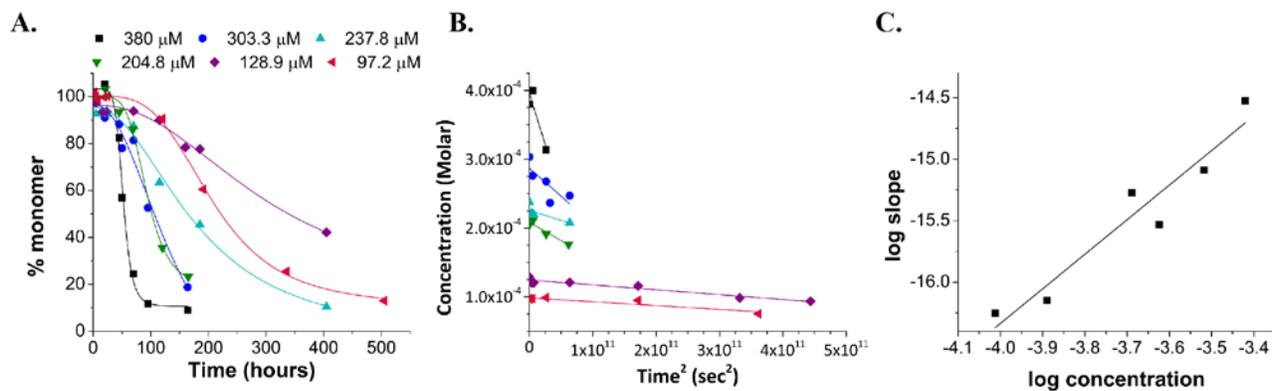


Figure A-4. K<sub>2</sub>Q<sub>9</sub>P<sup>D</sup>Q<sub>9</sub>K<sub>2</sub> kinetics, time<sup>2</sup>, and log-log plot.

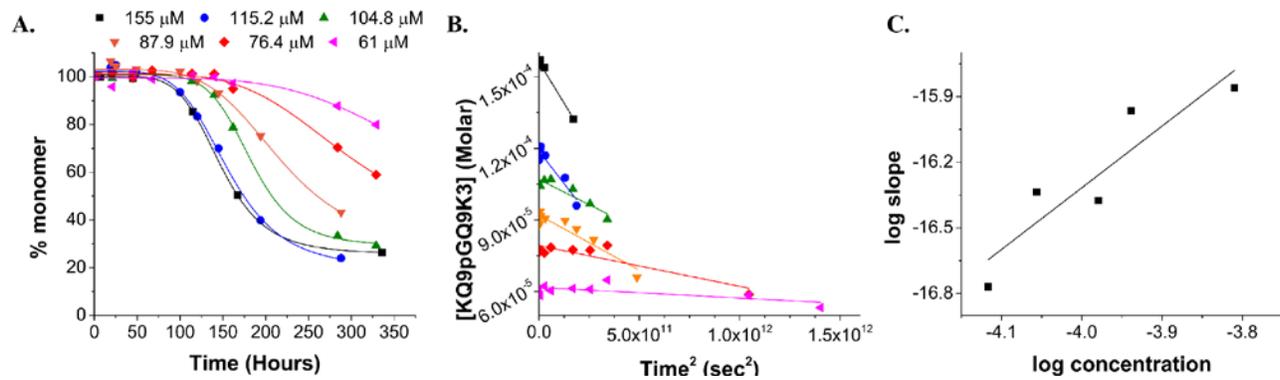


Figure A-5. KQ<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K<sub>3</sub> kinetics, time<sup>2</sup>, and log-log plot.

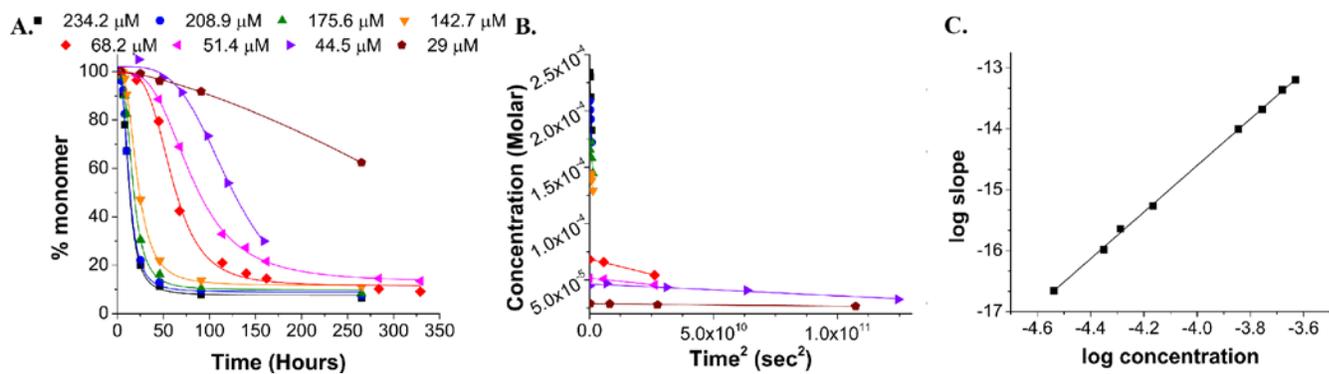


Figure A-6. K Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K kinetics, time<sup>2</sup>, and log-log plot.

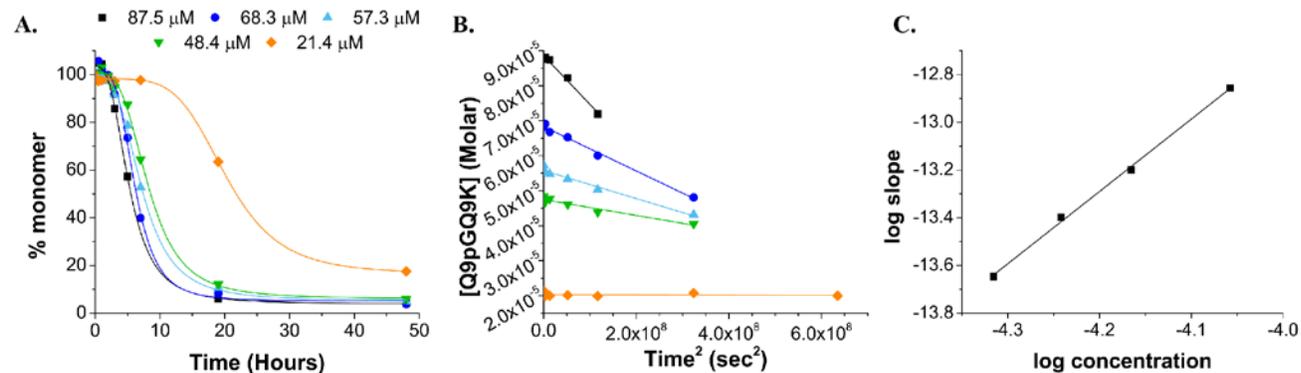


Figure A-7. Q<sub>9</sub>P<sup>D</sup>GQ<sub>9</sub>K kinetics, time<sup>2</sup>, and log-log plot.

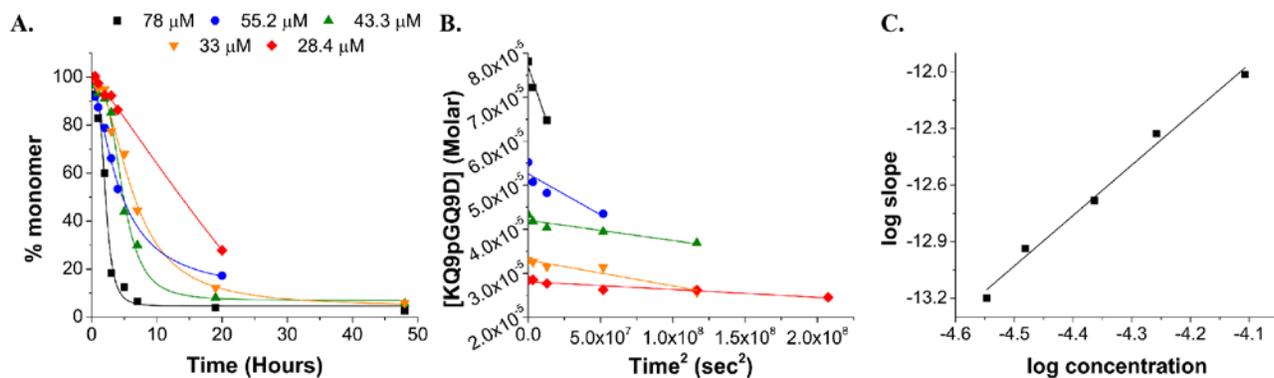


Figure A-8.  $KQ_9P^D GQ_9D$  kinetics,  $time^2$ , and log-log plot.

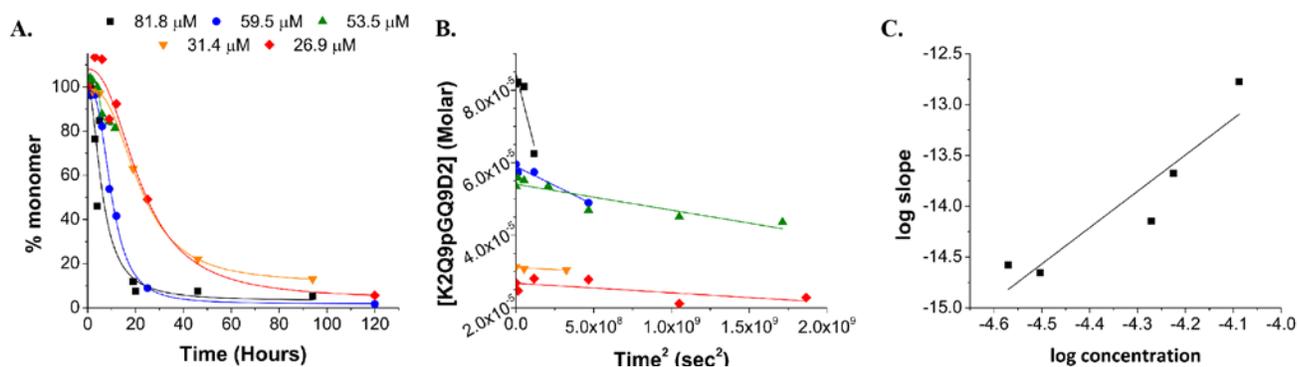
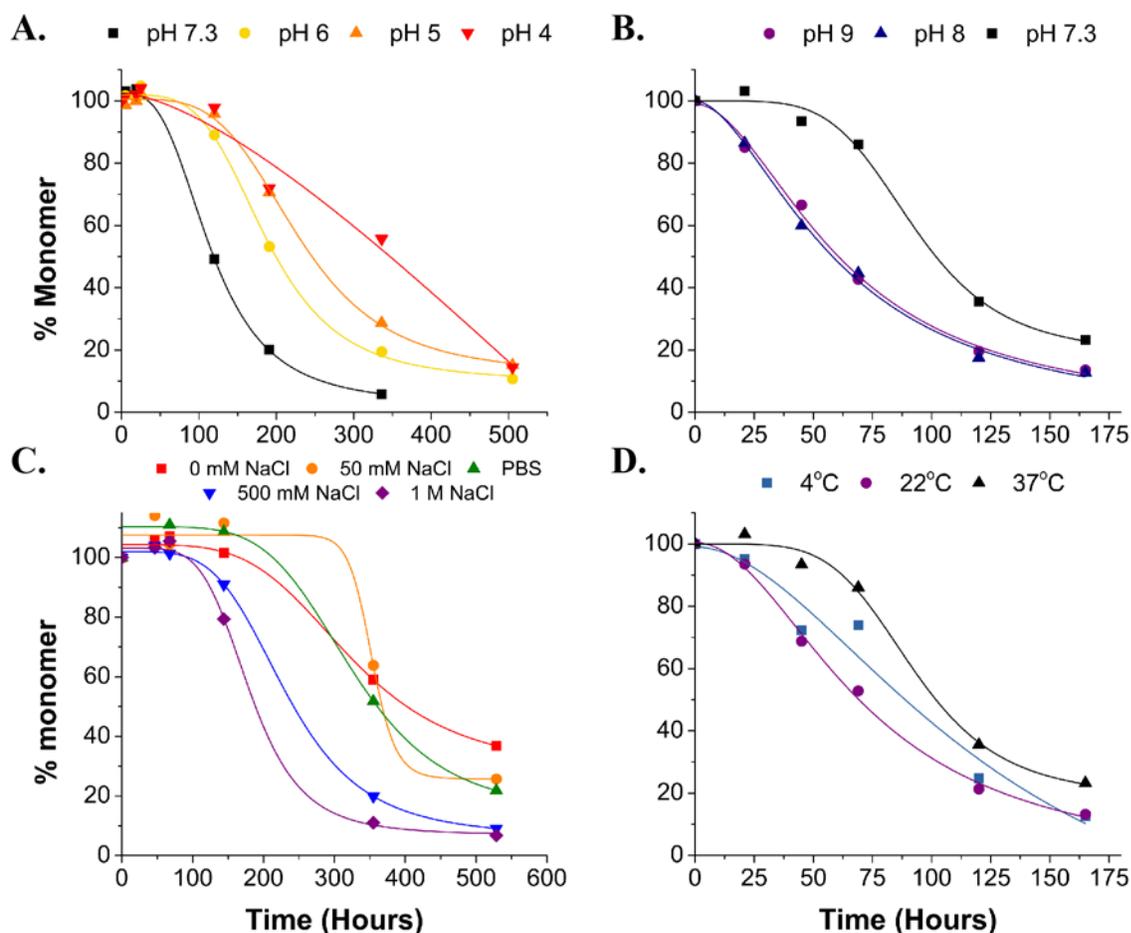


Figure A-9.  $K_2Q_9P^D GQ_9D_2$  kinetics,  $time^2$ , and log-log plot.

## A.2 REACTION CONDITIONS' ROLE IN AGGREGATION

$K_2Q_9P^D GQ_9K_2$  was investigated under a range of reaction conditions including varied pH, salt (NaCl) concentration, and temperature in an effort to determine the role they had on aggregation.

The kinetics of these reactions are shown in **Figure A-10**.



**Figure A-10.** Reactions of  $K_2Q_9P^D GQ_9K_2$  at various pH, salt concentrations, and temperatures.

As pH is shifted from acidic to basic, aggregation rate of  $K_2Q_9P^D GQ_9K_2$  increases. The pI of this peptide is 10.48, so this enhancement of aggregation rate may correspond to a decrease in net charge of the molecule as it approaches the pH at which it will bear no charge. There is still a lag phase present at the beginning of the reaction, even at pH 9, suggesting that the decrease in net charge has not resulted in a complete loss of solubility. No concentration dependence was measured for any of these variable reaction conditions, so it is not clear whether neutralization of the  $K_2Q_9P^D GQ_9K_2$  net charge had any effect on aggregation mechanism, as with other peptides with no net charge described in Chapter 4.0. Consistent with the results in that chapter, a decrease in repulsion does correspond in an increase in aggregation rate, suggesting an enhancement in  $\beta$ -hairpin formation and fibril elongation as pH is increased.

High ionic concentrations would be expected to have a screening effect on charged regions of  $K_2Q_9P^D GQ_9K_2$ . **Figure A-10 C** demonstrates that increases in salt concentration (NaCl, specifically) results in enhancement of aggregation rates, again suggesting that decreases in repulsion encourage spontaneous aggregation.

Finally, reactions at different temperatures also demonstrate differences in aggregation rates. These differences are subtle, but low temperatures appear to enhance aggregation kinetics. This trend is not due to any change in perceived charge, and has been observed previously.

### A.3 COMPLETE KINETICS DATA FOR VARIABLE REPEAT POLYQ

The kinetics,  $time^2$  and log-log plots for the variable repeat length peptides not shown in Chapter 5.0, Section 5.2.2. For all peptides, kinetic traces are shown in (A), the  $time^2$  plots for the early phase kinetics in (B), and the resulting log-log plots in (C). Only peptides with good fits ( $R^2 \geq 0.7$ ) to the  $time^2$  plot were used in log-log plots.

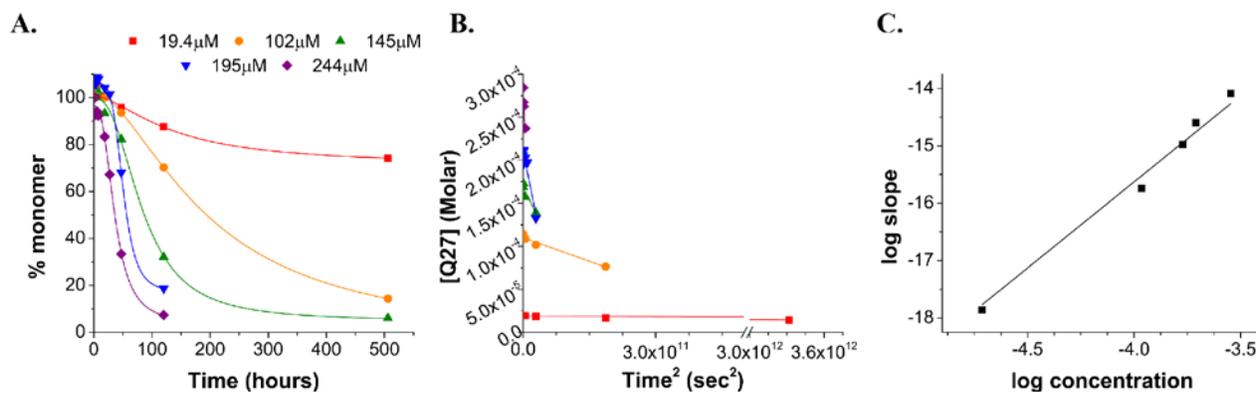


Figure A-11.  $K_2Q_{27}K_2$  kinetics,  $time^2$ , and log-log plot.

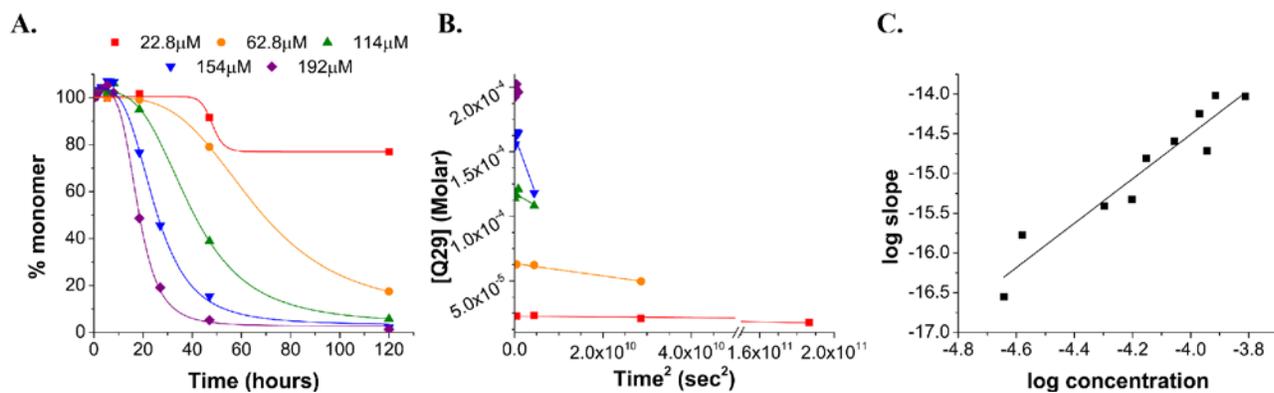


Figure A-12.  $K_2Q_{29}K_2$  kinetics, time<sup>2</sup>, and log-log plot.

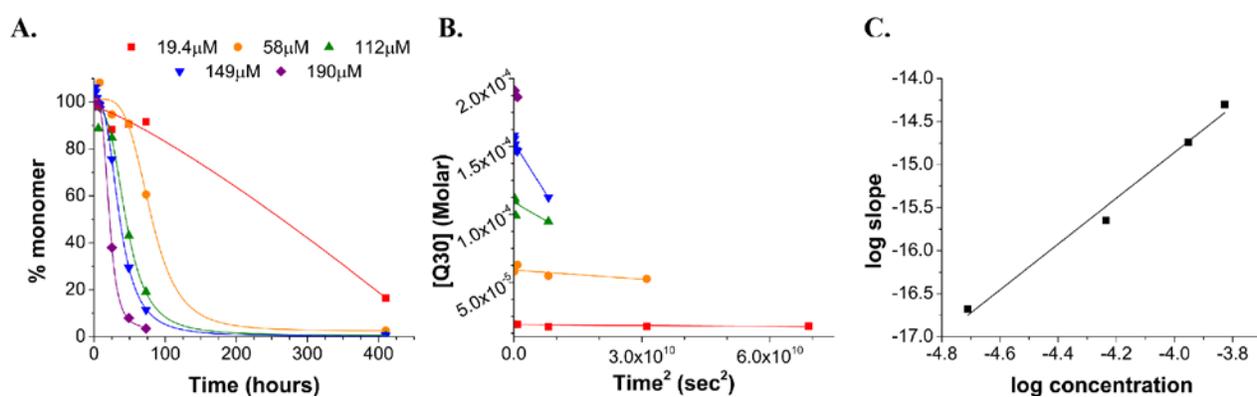


Figure A-13.  $K_2Q_{30}K_2$  kinetics, time<sup>2</sup>, and log-log plot.

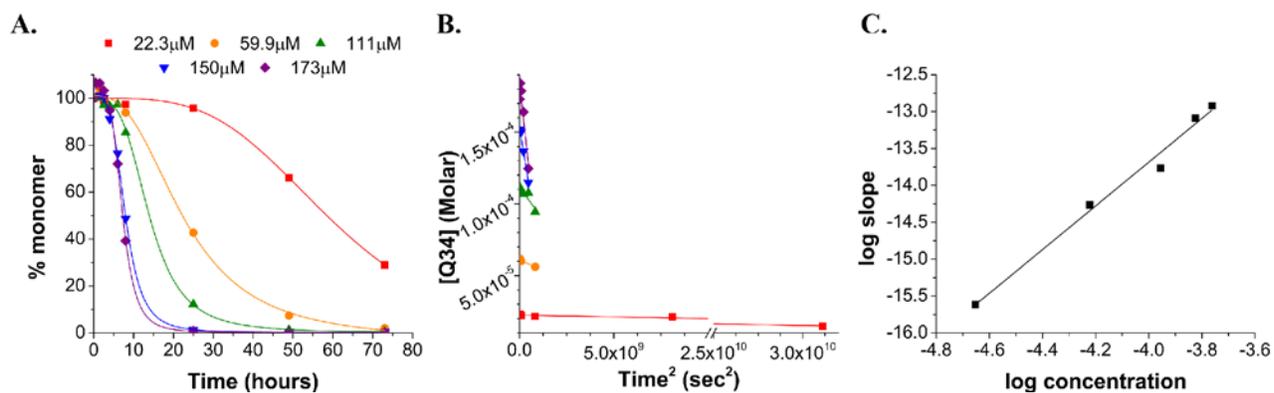


Figure A-14.  $K_2Q_{34}K_2$  kinetics, time<sup>2</sup>, and log-log plot.

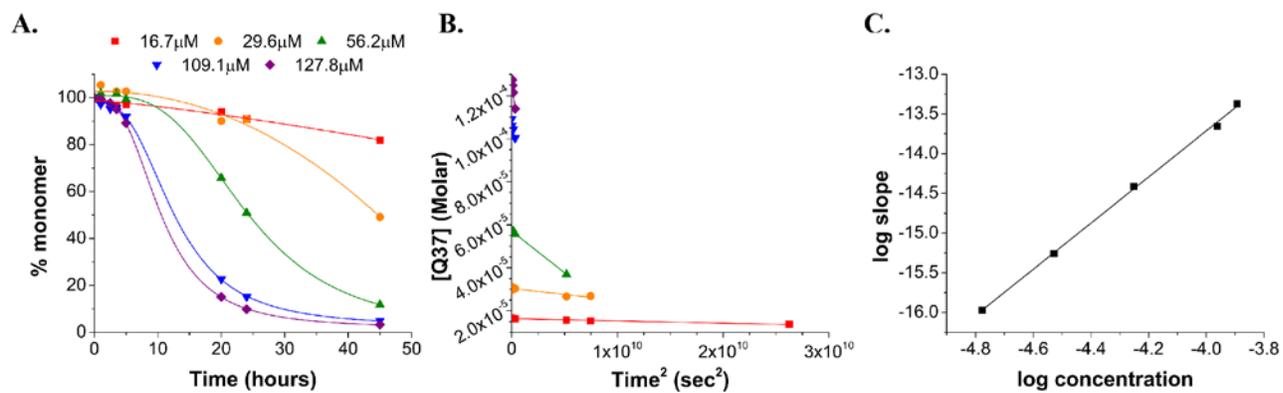


Figure A-15.  $K_2Q_{37}K_2$  kinetics,  $\text{time}^2$ , and log-log plot.

## BIBLIOGRAPHY

1. Wetzel, R., S. Shivaprasad, and A.D. Williams, *Plasticity of amyloid fibrils*. *Biochemistry*, 2007. **46**(1): p. 1-10.
2. Louis, J.M., et al., *The GBI Amyloid Fibril: Recruitment of the Peripheral  $\beta$ -Strands of the Domain Swapped Dimer into the Polymeric Interface*. *Journal of Molecular Biology*, 2005. **348**(3): p. 687-698.
3. Dobson, C.M., *Protein misfolding, evolution and disease*. *Trends Biochem Sci*, 1999. **24**(9): p. 329-32.
4. Dobson, C.M., *Principles of protein folding, misfolding and aggregation*. *Semin Cell Dev Biol*, 2004. **15**(1): p. 3-16.
5. Chiti, F. and C.M. Dobson, *Amyloid formation by globular proteins under native conditions*. *Nat Chem Biol*, 2009. **5**(1): p. 15-22.
6. Dobson, C.M., *The structural basis of protein folding and its links with human disease*. *Philos Trans R Soc Lond B Biol Sci*, 2001. **356**(1406): p. 133-45.
7. Maji, S.K., et al., *Functional amyloids as natural storage of peptide hormones in pituitary secretory granules*. *Science*, 2009. **325**(5938): p. 328-32.
8. Narayanaswamy, R., et al., *Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation*. *Proc Natl Acad Sci U S A*, 2009. **106**(25): p. 10147-52.
9. Maury, C.P.J., *The emerging concept of functional amyloid*. *Journal of Internal Medicine*, 2009. **265**(3): p. 329-334.
10. Sunde, M. and C. Blake, *The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction*, in *Advances in Protein Chemistry*, D.S.E. Frederic M. Richards and S.K. Peter, Editors. 1997, Academic Press. p. 123-159.
11. Toyama, B.H. and J.S. Weissman, *Amyloid Structure: Conformational Diversity and Consequences*. *Annual Review of Biochemistry*, 2011. **80**(1): p. 557-585.
12. Margittai, M. and R. Langen, *Fibrils with parallel in-register structure constitute a major class of amyloid fibrils: molecular insights from electron paramagnetic resonance spectroscopy*. *Quarterly Reviews of Biophysics*, 2008. **41**(3-4): p. 265-297.
13. Wetzel, R., *Ideas of Order for Amyloid Fibril Structure*. *Structure*, 2002. **10**(8): p. 1031-1036.
14. Kajava, A.V., U. Baxa, and A.C. Steven, *Beta arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils*. *FASEB J*, 2010. **24**(5): p. 1311-9.

15. Schneider, R., et al., *Structural Characterization of Polyglutamine Fibrils by Solid-State NMR Spectroscopy*. Journal of Molecular Biology, 2011. **412**(1): p. 121-136.
16. Sibanda, B.L. and J.M. Thornton, *[beta]-Hairpin families in globular proteins*. Nature, 1985. **316**(6024): p. 170-174.
17. Milner-White, E.J. and R. Poet, *Four classes of beta-hairpins in proteins*. Biochem J, 1986. **240**(1): p. 289-92.
18. Kar, K., et al., *Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent*. Nat Struct Mol Biol, 2011. **18**(3): p. 328-336.
19. Kar, K., et al.,  *$\beta$ -Hairpin-Mediated Nucleation of Polyglutamine Amyloid Formation*. Journal of Molecular Biology, 2013. **425**(7): p. 1183-97.
20. Sikorski, P. and E. Atkins, *New Model for Crystalline Polyglutamine Assemblies and Their Connection with Amyloid Fibrils*. Biomacromolecules, 2004. **6**(1): p. 425-432.
21. Harper, J.D., C.M. Lieber, and P.T. Lansbury Jr, *Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid- $\beta$  protein*. Chemistry & Biology, 1997. **4**(12): p. 951-959.
22. Ferrone, F., [17] *Analysis of protein aggregation kinetics*, in *Methods in Enzymology*, W. Ronald, Editor. 1999, Academic Press. p. 256-274.
23. Knowles, T.P.J., et al., *An Analytical Solution to the Kinetics of Breakable Filament Assembly*. Science, 2009. **326**(5959): p. 1533-1537.
24. Cohen, S.I.A., et al., *Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments*. The Journal of Chemical Physics, 2011. **135**(6): p. -.
25. Williams, A.D., S. Shivaprasad, and R. Wetzel, *Alanine Scanning Mutagenesis of A $\beta$ (1-40) Amyloid Fibril Stability*. Journal of Molecular Biology, 2006. **357**(4): p. 1283-1294.
26. Bitan, G., A. Lomakin, and D.B. Teplow, *Amyloid  $\beta$ -Protein Oligomerization: PRENUCLEATION INTERACTIONS REVEALED BY PHOTO-INDUCED CROSS-LINKING OF UNMODIFIED PROTEINS*. Journal of Biological Chemistry, 2001. **276**(37): p. 35176-35184.
27. Walsh, D.M., et al., *Amyloid  $\beta$ -Protein Fibrillogenesis: DETECTION OF A PROTOFIBRILLAR INTERMEDIATE*. Journal of Biological Chemistry, 1997. **272**(35): p. 22364-22372.
28. Bitan, G., et al., *Amyloid  $\beta$ -protein (A $\beta$ ) assembly: A $\beta$ 40 and A $\beta$ 42 oligomerize through distinct pathways*. Proceedings of the National Academy of Sciences, 2003. **100**(1): p. 330-335.
29. Haass, C. and D.J. Selkoe, *Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid [beta]-peptide*. Nat Rev Mol Cell Biol, 2007. **8**(2): p. 101-112.
30. Chiti, F. and C.M. Dobson, *Protein Misfolding, Functional Amyloid, and Human Disease*. Annual Review of Biochemistry, 2006. **75**(1): p. 333-366.
31. Modler, A.J., et al., *Assembly of Amyloid Protofibrils via Critical Oligomers—A Novel Pathway of Amyloid Formation*. Journal of Molecular Biology, 2003. **325**(1): p. 135-148.
32. Gosal, W.S., et al., *Competing Pathways Determine Fibril Morphology in the Self-assembly of  $\beta$ 2-Microglobulin into Amyloid*. Journal of Molecular Biology, 2005. **351**(4): p. 850-864.
33. Walsh, D.M., et al., *Amyloid  $\beta$ -Protein Fibrillogenesis: STRUCTURE AND BIOLOGICAL ACTIVITY OF PROTOFIBRILLAR INTERMEDIATES*. Journal of Biological Chemistry, 1999. **274**(36): p. 25945-25952.

34. Serio, T.R., et al., *Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant*. *Science*, 2000. **289**(5483): p. 1317-1321.
35. Cannon, M.J., et al., *Kinetic analysis of beta-amyloid fibril elongation*. *Analytical Biochemistry*, 2004. **328**(1): p. 67-75.
36. Esler, W.P., et al., *Alzheimer's Disease Amyloid Propagation by a Template-Dependent Dock-Lock Mechanism*. *Biochemistry*, 2000. **39**(21): p. 6288-6295.
37. O'Nuallain, B., et al., *Thermodynamics of A $\beta$ (1-40) Amyloid Fibril Elongation*. *Biochemistry*, 2005. **44**(38): p. 12709-12718.
38. Paulson, H.L. and K.H. Fischbeck, *Trinucleotide repeats in neurogenetic disorders*. *Annu Rev Neurosci*, 1996. **19**: p. 79-107.
39. Liu, Y. and S.H. Wilson, *DNA base excision repair: a mechanism of trinucleotide repeat expansion*. *Trends Biochem Sci*, 2012. **37**(4): p. 162-72.
40. Orr, H.T. and H.Y. Zoghbi, *Trinucleotide Repeat Disorders*. *Annual Review of Neuroscience*, 2007. **30**(1): p. 575-621.
41. Bates, G.P.B., C, *The Polyglutamine Diseases*, in *Huntington's Disease*, G.P. Bates, Harper, P .S. & Jones, L., Editor. 2002, Oxford University Press: Oxford, U.K. p. 429-472.
42. Wilburn, B., et al., *An Antisense CAG Repeat Transcript at JPH3 Locus Mediates Expanded Polyglutamine Protein Toxicity in Huntington's Disease-like 2 Mice*. *Neuron*, 2011. **70**(3): p. 427-440.
43. Cummings, C.J. and H.Y. Zoghbi, *Fourteen and counting: unraveling trinucleotide repeat diseases*. *Human Molecular Genetics*, 2000. **9**(6): p. 909-916.
44. Margolis, R.L., D.D. Rudnicki, and S.E. Holmes, *Huntington's disease like-2: review and update*. *Acta Neurol Taiwan*, 2005. **14**(1): p. 1-8.
45. Lin, X., C.J. Cummings, and H.Y. Zoghbi, *Expanding Our Understanding of Polyglutamine Diseases through Mouse Models*. *Neuron*, 1999. **24**(3): p. 499-502.
46. Zoghbi, H.Y. and H.T. Orr, *Polyglutamine diseases: protein cleavage and aggregation*. *Curr Opin Neurobiol*, 1999. **9**(5): p. 566-70.
47. Shizuka, M., et al., *Molecular analysis of a de novo mutation for spinocerebellar ataxia type 6 and (CAG) $n$  repeat units in normal elder controls*. *Journal of the Neurological Sciences*, 1998. **161**(1): p. 85-87.
48. Yabe, I., et al., *SCA6 mutation analysis in a large cohort of the Japanese patients with late-onset pure cerebellar ataxia*. *Journal of the Neurological Sciences*, 1998. **156**(1): p. 89-95.
49. Duyao, M., et al., *Trinucleotide repeat length instability and age of onset in Huntington's disease*. *Nat Genet*, 1993. **4**(4): p. 387-392.
50. Gusella, J.F. and M.E. MacDonald, *Molecular genetics: Unmasking polyglutamine triggers in neurodegenerative disease*. *Nat Rev Neurosci*, 2000. **1**(2): p. 109-115.
51. Bhattacharyya, A., *Oligoproline effects on polyglutamine conformation and aggregation*. *J. Mol. Biol.*, 2006. **355**: p. 524-535.
52. Darnell, G., et al., *Flanking Polyproline Sequences Inhibit  $\beta$ -Sheet Structure in Polyglutamine Segments by Inducing PPII-like Helix Structure*. *Journal of Molecular Biology*, 2007. **374**(3): p. 688-704.
53. Zoghbi, H.Y. and H.T. Orr, *Glutamine repeats and neurodegeneration*. *Annu Rev Neurosci*, 2000. **23**: p. 217-47.

54. Karlin, S. and C. Burge, *Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development*. Proceedings of the National Academy of Sciences, 1996. **93**(4): p. 1560-1565.
55. Gatchel, J.R. and H.Y. Zoghbi, *Diseases of Unstable Repeat Expansion: Mechanisms and Common Principles*. Nat Rev Genet, 2005. **6**(10): p. 743-755.
56. Gu, X., et al., *Serines 13 and 16 Are Critical Determinants of Full-Length Human Mutant Huntingtin Induced Disease Pathogenesis in HD Mice*. Neuron, 2009. **64**(6): p. 828-840.
57. Chen, H.K., et al., *Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1*. Cell, 2003. **113**(4): p. 457-68.
58. Emamian, E.S., et al., *Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in SCA1 transgenic mice*. Neuron, 2003. **38**(3): p. 375-87.
59. Steffan, J.S., et al., *SUMO modification of Huntingtin and Huntington's disease pathology*. Science, 2004. **304**(5667): p. 100-4.
60. Mishra, R., et al., *Serine phosphorylation suppresses huntingtin amyloid accumulation by altering protein aggregation properties*. J Mol Biol, 2012. **424**(1-2): p. 1-14.
61. Ross, C.A., *Polyglutamine Pathogenesis: Emergence of Unifying Mechanisms for Huntington's Disease and Related Disorders*. Neuron, 2002. **35**(5): p. 819-822.
62. Perutz, M.F., *Glutamine repeats and neurodegenerative diseases: molecular aspects*. Trends in Biochemical Sciences, 1999. **24**(2): p. 58-63.
63. Scherzinger, E., et al., *Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology*. Proc Natl Acad Sci U S A, 1999. **96**(8): p. 4604-9.
64. Chen, S., et al., *Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity*. Journal of Molecular Biology, 2001. **311**(1): p. 173-182.
65. Chen, S., et al., *Amyloid-like Features of Polyglutamine Aggregates and Their Assembly Kinetics†*. Biochemistry, 2002. **41**(23): p. 7391-7399.
66. Pang, J.T., et al., *Neuronal intranuclear inclusions in SCA2: a genetic, morphological and immunohistochemical study of two cases*. Brain, 2002. **125**(3): p. 656-663.
67. Glickman, M.H. and A. Ciechanover, *The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction*. Physiological Reviews, 2002. **82**(2): p. 373-428.
68. Ross, C.A., *Intranuclear Neuronal Inclusions: A Common Pathogenic Mechanism for Glutamine-Repeat Neurodegenerative Diseases?* Neuron, 1997. **19**(6): p. 1147-1150.
69. Yang, W., et al., *Aggregated polyglutamine peptides delivered to nuclei are toxic to mammalian cells*. Hum Mol Genet, 2002. **11**(23): p. 2905-17.
70. Arrasate, M., et al., *Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death*. Nature, 2004. **431**: p. 805-810.
71. Sahl, S.J., et al., *Cellular Inclusion Bodies of Mutant Huntingtin Exon 1 Obscure Small Fibrillar Aggregate Species*. Sci. Rep., 2012. **2**.
72. Bemporad, F. and F. Chiti, *Pathways of Amyloid Formation*, in *Amyloid Fibrils and Prefibrillar Aggregates*. 2013, Wiley-VCH Verlag GmbH & Co. KGaA. p. 151-166.
73. Jayaraman, M., R. Kodali, and R. Wetzel, *The impact of ataxin-1-like histidine insertions on polyglutamine aggregation*. Protein Engineering Design and Selection, 2009. **22**(8): p. 469-478.

74. Chen, S., F.A. Ferrone, and R. Wetzel, *Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation*. Proceedings of the National Academy of Sciences, 2002. **99**(18): p. 11884-11889.
75. Cohen, S.I.A., et al., *The Kinetics and Mechanisms of Amyloid Formation*, in *Amyloid Fibrils and Prefibrillar Aggregates*. 2013, Wiley-VCH Verlag GmbH & Co. KGaA. p. 183-209.
76. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions*. Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
77. Altschuler, E.L., et al., *Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides*. J Pept Res, 1997. **50**(1): p. 73-5.
78. Klein, F.A., et al., *Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient*. J Mol Biol, 2007. **371**(1): p. 235-44.
79. Crick, S.L., et al., *Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions*. Proceedings of the National Academy of Sciences, 2006. **103**(45): p. 16764-16769.
80. Walters, R.H. and R.M. Murphy, *Examining Polyglutamine Peptide Length: A Connection between Collapsed Conformations and Increased Aggregation*. Journal of Molecular Biology, 2009. **393**(4): p. 978-992.
81. Digambaranath, J.L., et al., *An accurate model of polyglutamine*. Proteins: Structure, Function, and Bioinformatics, 2011. **79**(5): p. 1427-1440.
82. Masino, L., et al., *Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins*. FEBS Lett, 2002. **513**(2-3): p. 267-72.
83. Wang, X., et al., *Characterizing the conformational ensemble of monomeric polyglutamine*. Proteins: Structure, Function, and Bioinformatics, 2006. **63**(2): p. 297-311.
84. Baldwin, R.L., *Energetics of Protein Folding*. Journal of Molecular Biology, 2007. **371**(2): p. 283-301.
85. Bhattacharyya, A.M., A.K. Thakur, and R. Wetzel, *Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15400-15405.
86. Persichetti, F., et al., *Mutant Huntingtin Forms in Vivo Complexes with Distinct Context-Dependent Conformations of the Polyglutamine Segment*. Neurobiology of Disease, 1999. **6**(5): p. 364-375.
87. Nagai, Y., et al., *A toxic monomeric conformer of the polyglutamine protein*. Nat Struct Mol Biol, 2007. **14**(4): p. 332-340.
88. Peters-Libeu, C., et al., *Disease-Associated Polyglutamine Stretches in Monomeric Huntingtin Adopt a Compact Structure*. Journal of Molecular Biology, 2012. **421**(4-5): p. 587-600.
89. Schaffar, G., *Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation*. Mol. Cell, 2004. **15**: p. 95-105.
90. Zhang, Q.C., et al., *A compact beta model of huntingtin toxicity*. J Biol Chem, 2011. **286**(10): p. 8188-96.
91. Bennett, M.J., et al., *A linear lattice model for polyglutamine in CAG-expansion diseases*. Proceedings of the National Academy of Sciences, 2002. **99**(18): p. 11634-11639.

92. Klein, F.A.C., et al., *Linear and extended: a common polyglutamine conformation recognized by the three antibodies MW1, 1C2 and 3B5H10*. Human Molecular Genetics, 2013. **22**(20): p. 4215-4223.
93. Bulone, D., et al., *The Interplay between PolyQ and Protein Context Delays Aggregation by Forming a Reservoir of Protofibrils*. PLoS ONE, 2006. **1**(1): p. e111.
94. Lee, C.C., R.H. Walters, and R.M. Murphy, *Reconsidering the Mechanism of Polyglutamine Peptide Aggregation*†. Biochemistry, 2007. **46**(44): p. 12810-12820.
95. Demuro, A., et al., *Calcium dysregulation and membrane disruption as a ubiquitous neurotoxic mechanism of soluble amyloid oligomers*. J Biol Chem, 2005. **280**(17): p. 17294-300.
96. Thakur, A.K., et al., *Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism*. Nat Struct Mol Biol, 2009. **16**(4): p. 380-389.
97. Sharma, D., et al., *Polyglutamine homopolymers having 8–45 residues form slablike  $\beta$ -crystallite assemblies*. Proteins: Structure, Function, and Bioinformatics, 2005. **61**(2): p. 398-411.
98. Sivanandam, V.N., et al., *The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils*. J Am Chem Soc, 2011. **133**(12): p. 4558-66.
99. Thakur, A.K. and R. Wetzel, *Mutational analysis of the structural organization of polyglutamine aggregates*. Proceedings of the National Academy of Sciences, 2002. **99**(26): p. 17014-17019.
100. Venkatraman, J., S.C. Shankaramma, and P. Balaram, *Design of folded peptides*. Chem Rev, 2001. **101**(10): p. 3131-52.
101. Poirier, M.A., H. Jiang, and C.A. Ross, *A structure-based analysis of huntingtin mutant polyglutamine aggregation and toxicity: evidence for a compact beta-sheet structure*. Human Molecular Genetics, 2005. **14**(6): p. 765-774.
102. Ramirez-Alvarado, M., F.J. Blanco, and L. Serrano, *De novo design and structural analysis of a model beta-hairpin peptide system*. Nat Struct Biol, 1996. **3**(7): p. 604-12.
103. Cochran, A.G., N.J. Skelton, and M.A. Starovasnik, *Tryptophan zippers: stable, monomeric beta-hairpins*. Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5578-83.
104. Hughes, R.M. and M.L. Waters, *Model systems for beta-hairpins and beta-sheets*. Curr Opin Struct Biol, 2006. **16**(4): p. 514-24.
105. Fesinmeyer, R.M., F.M. Hudson, and N.H. Andersen, *Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin*. J Am Chem Soc, 2004. **126**(23): p. 7238-43.
106. Santiveri, C.M., et al., *Context-dependence of the contribution of disulfide bonds to beta-hairpin stability*. Chemistry, 2008. **14**(2): p. 488-99.
107. Chen, S. and R. Wetzel, *Solubilization and disaggregation of polyglutamine peptides*. Protein Science, 2001. **10**(4): p. 887-891.
108. Perutz, M.F., et al., *Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases*. Proc Natl Acad Sci U S A, 1994. **91**(12): p. 5355-8.
109. Kato, A., et al., *Mutational analysis of protein solubility enhancement using short peptide tags*. Biopolymers, 2007. **85**(1): p. 12-8.
110. Sharma, D., et al., *Peptide models for inherited neurodegenerative disorders: conformation and aggregation properties of long polyglutamine peptides with and without interruptions*. FEBS Letters, 1999. **456**(1): p. 181-185.

111. Altschuler, E.L., et al., *Structure of polyglutamine*. FEBS Letters, 2000. **472**(1): p. 166-167.
112. Brahmachari, S.K., et al., *Probing the polyglutamine puzzle in neurological disorders*. FEBS Letters, 2000. **472**(1): p. 167-168.
113. Ellisdon, A.M., B. Thomas, and S.P. Bottomley, *The Two-stage Pathway of Ataxin-3 Fibrillogenesis Involves a Polyglutamine-independent Step*. Journal of Biological Chemistry, 2006. **281**(25): p. 16888-16896.
114. Scherzinger, E., et al., *Huntingtin-Encoded Polyglutamine Expansions Form Amyloid-like Protein Aggregates In Vitro and In Vivo*. Cell, 1997. **90**(3): p. 549-558.
115. Kurouski, D., et al., *Levels of supramolecular chirality of polyglutamine aggregates revealed by vibrational circular dichroism*. FEBS Lett, 2013. **587**(11): p. 1638-43.
116. Eaton, W.A. and J. Hofrichter, *Sickle cell hemoglobin polymerization*. Adv Protein Chem, 1990. **40**: p. 63-279.
117. Ferrone, F.A., *Nucleation: the connections between equilibrium and kinetic behavior*. Methods Enzymol, 2006. **412**: p. 285-99.
118. Collins, S.R., et al., *Mechanism of prion propagation: amyloid growth occurs by monomer addition*. PLoS Biol, 2004. **2**(10): p. e321.
119. Tanaka, M., et al., *The physical basis of how prion conformations determine strain phenotypes*. Nature, 2006. **442**(7102): p. 585-589.
120. Vitalis, A. and R.V. Pappu, *Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides*. Biophysical Chemistry, 2011. **159**(1): p. 14-23.
121. Vitalis, A., N. Lyle, and R.V. Pappu, *Thermodynamics of  $\beta$ -Sheet Formation in Polyglutamine*. Biophysical Journal, 2009. **97**(1): p. 303-311.
122. Lyle, N.J., S.L. Crick, and R.V. Pappu, *Alterations to the Conformational Ensemble and Intermolecular Associations of Polyglutamine Due to Charged Side Chains at the N- and C-Termini*. Biophysical Journal, 2011. **100**(3, Supplement 1): p. 63a.
123. Lakhani, V.V., F. Ding, and N.V. Dokholyan, *Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences*. PLoS Comput Biol, 2010. **6**(4): p. e1000772.
124. Williamson, T.E., et al., *Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin*. J Mol Biol, 2010. **396**(5): p. 1295-309.
125. O'Nuallain, B., et al., *Kinetics and Thermodynamics of Amyloid Assembly Using a High-Performance Liquid Chromatography-Based Sedimentation Assay*, in *Methods in Enzymology*, K. Indu and W. Ronald, Editors. 2006, Academic Press. p. 34-74.
126. Jayaraman, M., et al., *Assays for studying nucleated aggregation of polyglutamine proteins*. Methods, 2011. **53**(3): p. 246-254.
127. Kar, K., et al., *D-polyglutamine amyloid recruits L-polyglutamine monomers and kills cells*. Journal of Molecular Biology, (0).
128. Jayaraman, M., et al., *Slow Amyloid Nucleation via  $\alpha$ -Helix-Rich Oligomeric Intermediates in Short Polyglutamine-Containing Huntingtin Fragments*. Journal of Molecular Biology, 2012. **415**(5): p. 881-899.
129. Kuipers, B.J.H. and H. Gruppen, *Prediction of Molar Extinction Coefficients of Proteins and Peptides Using UV Absorption of the Constituent Amino Acids at 214 nm To Enable Quantitative Reverse Phase High-Performance Liquid Chromatography-Mass*

- Spectrometry Analysis*. Journal of Agricultural and Food Chemistry, 2007. **55**(14): p. 5445-5451.
130. Diamandis, E.P., *Immunoassays with time-resolved fluorescence spectroscopy: Principles and applications*. Clinical Biochemistry, 1988. **21**(3): p. 139-150.
  131. O'Nuallain, B., et al., *Seeding Specificity in Amyloid Growth Induced by Heterologous Fibrils*. Journal of Biological Chemistry, 2004. **279**(17): p. 17490-17499.
  132. Cohen, S.I.A., et al., *From Macroscopic Measurements to Microscopic Mechanisms of Protein Aggregation*. Journal of Molecular Biology, 2012. **421**(2-3): p. 160-171.
  133. Fitzpatrick, A.W.P., et al., *Atomic structure and hierarchical assembly of a cross- $\beta$  amyloid fibril*. Proceedings of the National Academy of Sciences, 2013. **110**(14): p. 5468-5473.
  134. Carulla, N., et al., *Molecular recycling within amyloid fibrils*. Nature, 2005. **436**(7050): p. 554-558.
  135. Chen, S., et al., *Amyloid-like features of polyglutamine aggregates and their assembly kinetics*. Biochemistry, 2002. **41**: p. 7391-7399.
  136. Lee, A.C. and G.M. Crippen, *Predicting pKa*. Journal of Chemical Information and Modeling, 2009. **49**(9): p. 2013-2033.
  137. Grimsley, G.R., J.M. Scholtz, and C.N. Pace, *A summary of the measured pK values of the ionizable groups in folded proteins*. Protein Science, 2009. **18**(1): p. 247-251.
  138. Schlosshauer, M. and D. Baker, *Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness*. Protein Sci, 2004. **13**(6): p. 1660-9.
  139. Wetzel, R., *Physical Chemistry of Polyglutamine: Intriguing Tales of a Monotonous Sequence*. Journal of Molecular Biology, 2012. **421**(4-5): p. 466-490.