Washington University in St. Louis

Washington University Open Scholarship

January 2011

Inferring Aggregation Mechanisms Of Molecules Involved In Neurodegeneration Through Quantitative Studies Of Phase Behavior

Scott Crick
Washington University in St. Louis

Recommended Citation

Crick, Scott, "Inferring Aggregation Mechanisms Of Molecules Involved In Neurodegeneration Through Quantitative Studies Of Phase Behavior" (2011). *All Theses and Dissertations (ETDs)*. 76.
https://openscholarship.wustl.edu/etd/76

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science

Department of Biomedical Engineering

Dissertation Examination Committee:
Rohit V. Pappu, Chair
Marc Diamond
Donald Elbert
Elliot Elson
Carl Frieden
Eric Galburt
Jin-Moo Lee

Inferring Aggregation Mechanisms of Molecules Involved in Neurodegeneration through

Quantitative Studies of Phase Behavior

by

Scott L. Crick

A dissertation presented to
The Graduate School of Arts & Sciences of Washington University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August, 2011

Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Inferring aggregation mechanisms of molecules involved in neurodegeneration through

quantitative studies of phase behavior

by

Scott L. Crick

Doctor of Philosophy

Washington University in St. Louis, 2011

Research Advisors: Professor Rohit Pappu and Professor Carl Frieden

Polyglutamine is involved in at least nine known neurodegenerative diseases, the most prominent of which is Huntington's Disease. It is thought that polyglutamine aggregation leads to disease.

The biophysical mechanism of polyglutamine aggregation remains controversial as highlighted by conflicting proposals that have been put forth in the literature ranging from homogeneous nucleation to a more complex assembly mechanism that involves heterogeneous distributions of oligomers. Converging upon an accurate framework for describing polyglutamine aggregation *in vitro* is an essential first step for understanding how interactions in *cis* i.e., flanking sequences and *trans* i.e., heterotypic interactions in the cellular milieu shape self-assembly and the formation of inclusions.

In this work, we leverage concepts from polymer physics, to understand solution phase behavior of polyglutamine. Specifically, we first characterize water as poor solvent for polyglutamine. This classification suggests that polyglutamine forms collapsed structures in aqueous solution. At low concentrations, this will lead to homogeneously dispersed solutions of compact globules. At higher concentrations, the globules will coalesce leading to phase separation.

Next, we characterize the phase behavior of polyglutamine solutions and develop a reference phase diagram for polyglutamine peptides that provides thermodynamic constraints for aggregation mechanisms. Specifically, we measure temperature-dependent saturation concentrations of aqueous polyglutamine solutions containing 30 and 40 glutamine residues and either 2 or 4 flanking lysines. We used classical Flory-Huggins theory to construct the phase diagram for partitioning between soluble and insoluble phases from the measured saturation concentrations. The low-concentration arm of the phase diagram provides a thermodynamic basis for assessing aggregation propensity. For a given chain length, aggregation propensity increases as the number of lysine residues decrease highlighting the contributions from intermolecular electrostatic repulsions. For a fixed number of lysine residues, the aggregation propensity increases with increasing chain length, highlighting the intrinsic contributions of polyglutamine length to the driving forces for aggregation. The inferred phase diagrams provide thermodynamic constraints on the kinetic mechanisms for aggregation. In addition, at

physiological temperatures, the gap between the saturation curve and the instability boundary spans roughly two orders of magnitude. This suggests that the formation of metastable, higher-order clusters and conformational conversions within these clusters are likely precursors for polyglutamine aggregation thereby rationalizing a role for oligomers that have been observed in recent studies based on AFM and light scattering.

Finally, we apply our knowledge of the phase behavior of polyglutamine to understand mechanisms by which amyloid beta aggregation might be modulated by cellular activities. In particular, our experiments suggest that amyloid beta is taken up from the extracellular space by neurons, trafficked into acidic vesicles, and concentrated to levels known to support aggregation based on the phase diagram.

Acknowledgments

I would like to thank, first and foremost, my mentor, Rohit Pappu. He has supported me intellectually, financially, and emotionally to degrees that no reasonable person could expect. He has been a wonderful mentor, and has become a dear friend. I look forward to continuing our work together.

I would also like to thank my other mentor, Carl Frieden. Without Carl's guidance and expertise, I could have never accomplished all that I have in the past six years. Carl constantly challenged me with problems. If Carl didn't think I could solve them, he wouldn't have challenged me. When someone as accomplished as Carl believes in you, it makes you think you can do anything.

My parents and sister have supported me all my life. They are my number one fans and my closest allies. Without their love and encouragement, I am almost certain I could not have accomplished much of anything. They are my rock.

I would like to thank my girlfriend, Anna Bone, for all of her love and support, and, especially her patience over the past few months of writing my thesis. Anna has made me a better person, period. She is one of the most patient and kind people I have ever met, and she is starting to rub off on me. I live in my head too much of the time, and Anna keeps me grounded.

I would like to thank my lab mates. I have learned something from every one of them. In particular, I would like to thank Kanchan Garai, Nick Lyle, Albert Mao, Tim Williamson, and Andreas Vitalis. These people have taught me much, and I consider all of them to be more than my co-workers. They are my friends.

I would like to thank the members of my committee for reading and contributing to this work.

Scott L. Crick

Washington University in St. Louis
June 2011

v

Dedicated to my grandfather, Jack Luckett, a victim of Alzheimer's disease.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Preamble

The main focus of this thesis is on gaining an understanding of the aqueous solution phase behavior of polyglutamine, molecules comprised solely of repeating glutamine tracts. The behavior of this class of molecules in aqueous solution is much more complex than one would predict. The amino-acid glutamine is polar, and relatively soluble in water. (*1*) However, when concatenated into a homopeptide (from here on referred to as polyglutamine), the solubility decreases dramatically with increasing number of glutamines. (*2, 3*) Even heteropeptides which contain high percentages of glutamine residues are more insoluble than one would intuitively expect. (*2, 4-6*) Moreover, in aqueous solutions, polyglutamine is apparently devoid of structure prior to aggregation. Upon aggregation, polyglutamine can form many different types of structures, some of which are linear, worm-like chains rich in cross β-sheet structure. These linear aggregates are, by certain characteristics such as the binding of fluorescent dyes (*7*), amyloid-like. The aggregation of normally functional proteins into amyloid is thought to play a major role in many human diseases.(*8-11*) At least nine different proteins containing expanded tracts of polyglutamine are implicated in the development of neurodegenerative diseases in humans. (*12-19*) It is widely believed that the polyglutamine tracts are responsible for this pathological gain-of-function. (*20-29*)

Our major interest in polyglutamine is due to its involvement in neurodegenerative diseases, such as Huntington's Disease. This class of molecules is also interesting in its own right. That complex physicochemical behavior can arise from concatenation of a single, simple component is now more amazing to me near the end of my graduate career than it was at the beginning.

My work has focused on developing a quantitative understanding of polyglutamine aggregation using concepts borrowed from polymer physics, colloidal science, and solution phase thermodynamics. As a result, this thesis is organized in a way that reflects the synthesis of concepts from different fields. The first step was to obtain an understanding of the peptide in the monomeric (non-aggregated) form. What are the features of polyglutamine in solution? Why might it be prone to aggregate? The next step was to obtain an understanding of the protein aggregation process *in vitro*. What are the barriers to aggregation? What are the mechanisms by which aggregation can occur? When aggregation occurs, what species might be formed? What factors can modulate the aggregation? Finally, we attempted to move beyond well-controlled and stylized protein systems by investigating how aggregation of a protein associated with Alzheimer's disease can occur *in vivo* even though the solution phase thermodynamics seem to indicate that this is impossible. (*30-37*) What cellular factors and processes might play a role in modulating aggregation? How might aggregation lead to the development of disease?

This chapter is organized as follows: We start with a discussion of the relevance of polyglutamine in biological systems focusing mainly on its role in the development of several neurodegenerative diseases. We then provide an overview of studies focused on the biophysical characterization of polyglutamine. This includes a discussion of a commonly used framework for the study of aggregation-prone proteins. To orient the reader, we then provide a brief overview of the relevant concepts from polymer physics and solution phase thermodynamics which we use to guide our research. Finally, we provide a summary of this chapter and an outline for the rest of the thesis.

1.2 Relevance of Polyglutamine

1.2.1 Polyglutamine in disease

It is widely accepted that a wide variety of proteins can transition from normally functioning, soluble species into potentially deleterious aggregates, and that this accumulation of aggregates is associated with development of disease pathology. (*10, 38-43*) The formation of insoluble aggregates both *in vitro* and *in vivo* is a common characteristic of proteins implicated in a variety of diseases including systemic amyloidoses and neurodegenerative diseases including, but not limited to: Parkinson's disease, Alzheimer's disease, Huntington's disease, and prion diseases. (*9-11, 38-40, 42-48*) One class of the protein-aggregation related neurodegenerative diseases is known as polyglutamine expansion diseases because they all seem to be associated with the aggregation of proteins whose only common sequence feature is the presence of long polyglutamine tracts. (*12-28, 49-76*)

3

There are at least nine known polyglutamine expansion diseases (see Table 1.1) which include: Huntington's disease (HD), Spinal Bulbar Muscular Atrophy (SBMA), six different Spinocerebellar Ataxias (SCA), and Dentatorubral-Pallidoluysian Atrophy (DRPLA). (*12-19, 49-76*) In all of these diseases, the codon for the amino acid glutamine (CAG) is erroneously repeated in the coding region of the DNA, increasing the length of the polyglutamine in the resulting protein. The mechanism for this erroneous expansion of the CAG triplet during DNA replication is still not well understood. (*52, 77*)

Polyglutamine Disease	Host Protein	Average Molecular Weight (kDa)	Normal Repeat Length	Pathogenic Repeat Length
Spinobulbar Muscular Atrophy (SBMA)	Androgen Receptor	~99	9-36	38-62
Huntington's Disease (HD)	Huntingtin	~348	6-35	36-121
Spinocerebellar Ataxia Type 1 (SCA1)	Ataxin-1	~87	6-44	39-82
Spinocerebellar Ataxia Type 2 (SCA2)	Ataxin-2	~140	15-31	36-63
Spinocerebellar Ataxia Type 3 (SCA3)	Ataxin-3	~41	12-40	55-84
Spinocerebellar Ataxia Type 6 (SCA6)	α1A-voltage-dependent calcium channel subunit	~282	4-18	21-33
Spinocerebellar Ataxia Type 7 (SCA7)	Ataxin-7	~95	4-35	37-306
Spinocerebellar Ataxia Type 17 (SCA17)	TATA Binding Protein (TBP)	~38	25-42	45-63
Dentatorubral-pallidoluysian atrophy (DRPLA)	Atrophin-1	~125	6-35	49-88

Table 1.1 Overview of the polyglutamine expansion diseases. The nine polyglutamine expansion diseases are listed along with the host protein, the approximate average molecular weight taken from the NCBI Protein database, the normal polyglutamine repeat length, and the pathogenic polyglutamine repeat length.

Aside from containing uninterrupted stretches of polyglutamine, the host proteins involved in these diseases have very little else in common. The polyglutamine stretches can be of variable length, but there exists a fairly sharp threshold length for each protein above which the disease state emerges (see Table 1.1). (*12, 15-18, 51, 56, 59, 61, 65, 71*) As the length of the polyglutamine increases past the threshold length for a given host protein, the disease age-of-onset decreases (see Figure 1.1) and the disease severity at onset increases as well. (*12, 15-18, 51, 56, 59, 61, 65, 71*) This strong genotype-phenotype correlation suggests that polyglutamine plays a causative role in the development of disease.

As this thesis is concerned with the molecular aspects of polyglutamine expansion diseases and, in particular, of the role of polyglutamine itself, we will not provide a comprehensive review of the clinical features or genetics of these diseases. For a detailed treatment of these issues, we refer the interested reader to the following resources (*12, 50, 56, 78-83*). However, it is important to cover some of the molecular biology of these diseases to provide a cellular context for understanding processing of polyglutamine containing proteins and possible mechanisms for polyglutamine-mediated toxicity.
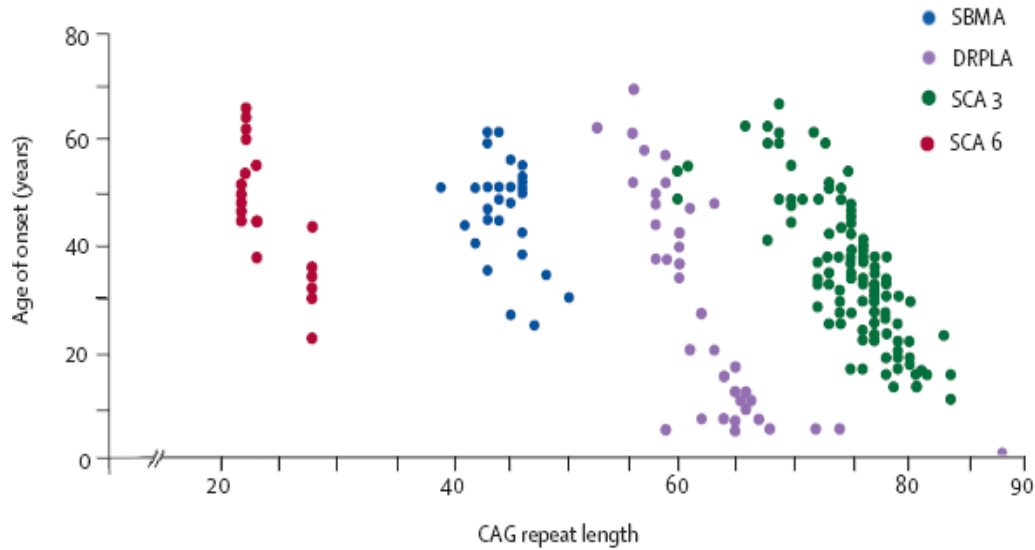
Figure 1.1 Age of onset versus CAG repeat length in four human polyglutamine expansion disorders. All polyglutamine expansion disorders display an inverse correlation between age of onset and CAG repeat length. However, the pathological length differs in each polyglutamine expansion disease (see also Table 1.1). (*84*)

Although there exists an inverse correlation between polyglutamine length in disease related proteins and the pathology of disease, this does not clearly distinguish polyglutamine as the cause. It could be possible that the repeated CAG codons in the transcribed DNA lead to the problem, or that the resulting RNA is somehow to blame. There is a class of diseases for which this appears to be the case, because there is a disease phenotype also associated with the length of the CAG repeat in the DNA, but the repeat occurs in non-coding regions of the disease gene. (*50, 52, 56, 57*) However, the length of the pathological repeats in these diseases is generally much longer, and the dysfunction typically affects many more systems than the polyglutamine expansion diseases. Both experimental and histopathological evidence strongly suggest a toxic

7

effect of polyglutamine containing protein aggregates in the nine diseases known as polyglutamine expansion diseases.

To establish a link between polyglutamine and pathology, it was necessary to assess if the host protein containing the pathological length of polyglutamine is expressed normally relative to the normal host protein. Using antibodies to regions at both ends of Huntingtin (the protein implicated in Huntington's disease), Jou and Myers demonstrated using Western blot analysis of cell lysates from human brain and both human and monkey cell lines that the full length Huntingtin was expressed normally even with a pathological length of polyglutamine. (*60*) Subsequently, Ikeda and coworkers expressed full-length ataxin-3 (the protein implicated in SCA3) containing both normal (Q_{35}) and pathogenic (Q_{79}) polyglutamine stretches in cultured COS cells. These cells were indistinguishable from cells expressing retinoic acid receptor as a control. When the researchers expressed only a fragment of ataxin-3 containing the polyglutamine region and the 42 amino acid C-terminus of ataxin-3 (referred to as $Q_{35}C$ and $Q_{79}C$), an abnormal phenotype developed exclusively in the cells expressing $Q_{79}C$. The phenotype was characterized by the development of punctuate aggregates and apoptosis. When only the polyglutamine segments were expressed, the abnormal phenotype was again seen only for cells expressing the pathological length of polyglutamine. The researchers also expressed $Q_{64}C$, where 64 is also a pathological length, and noted that cells expressing $Q_{64}C$ showed the abnormal phenotype, but to a lesser extent than cells expressing $Q_{79}C$. To test the hypothesis that it is the translated CAG responsible for the abnormal phenotype, the researchers added 79 CAG repeats just downstream of the coding region

8

for the construct encoding polyglutamine 35 and the C-terminal region of ataxin-3, and no abnormal phenotype was seen. Transgenic mice were created expressing full length ataxin-3 with Q_{35} or Q_{79} and the ataxin-3 fragments $Q_{35}C$ or $Q_{79}C$ in Purkinje cells, which are only mildly affected in SCA3 patients. Ataxic symptoms and histopathology were seen only in the $Q_{79}C$ mice. Interestingly, Burright and coworkers produced transgenic mice expressing full length ataxin-1 (the protein implicated in SCA1) also specifically in Purkinje cells which, in contrast to SCA3, are strongly affected in SCA1. They observed ataxic phenotypes only in mice expressing pathological lengths of polyglutamine. (*85*) The fact that full-length ataxin-1 containing a pathological length of polyglutamine caused disease phenotype in Purkinje cells (the primary cells affected in SCA1), whereas full-length ataxin-3 containing a pathological length of polyglutamine did not lead to disease phenotype in Purkinje cells (which are only mildly affected in SCA3), strongly suggests that proteolytic processing plays a major role in disease pathogenesis. Moreover, it is known that each polyglutamine disease differentially affects specific populations of neurons (*12, 18, 26, 49, 58*) even though the disease related proteins are ubiquitously expressed throughout the brain. Differences in proteolytic processing may be a major reason why neurons show differential susceptibility to the pathological forms of the disease related proteins. Importantly, it is known that the polyglutamine stretches are highly resistant to proteolysis (*86-88*), so regardless of which proteases the host protein is exposed to, the polyglutamine stretch is likely to remain intact.

A major histopathological feature of polyglutamine disease patients' brains is the presence of neuronal intranuclear inclusions. (*10, 13, 14, 16, 44, 89*) These intranuclear

inclusions are composed primarily of the host proteins and polyglutamine-containing proteolytic fragments. (*10, 13, 14, 16, 44, 89*) These proteins and fragments are often ubiquitinated (*14, 90*) suggesting that they were, at one time, targeted for degradation. The inclusions also tend to contain protein-processing machinery such as the 26S proteasome subunits and chaperones (*91*) which has led to a proposal that toxicity arises from disruption of protein folding homeostasis. (*92*) Numerous transcription factors (*93-95*) are also found in the inclusions, and this has led to an alternative hypothesis that transcriptional dysregulation is the source of cellular toxicity. (*96-98*)

There is direct evidence that polyglutamine aggregates are toxic to cells when delivered to the nucleus. (*99*) Yang et al. prepared aggregates *in vitro* of polyglutamine of the form $KKQ_{20}KK$ and $KKQ_{42}KK$ either with or without a nuclear localization signal (NLS). When the aggregates were added to cell culture, both COS-7 and PC-12 cells took up the aggregates. Aggregates of the polyglutamine peptides not containing the nuclear localization signal were sequestered in the cytoplasm and had no effect on cell viability. Aggregates of the polyglutamine peptides containing the NLS were visualized in the nucleus, and were toxic. The aggregates of the polyglutamine peptides containing only 20 glutamines were deemed to be just as toxic as the aggregates of the polyglutamine peptides containing 42 glutamines. As a control, aggregates of the peptide CspB-1 containing an NLS were also added to the cells, and they appeared to have no effect on the cell viability, even when present in the nucleus. These results suggest that aggregates of polyglutamine and polyglutamine containing proteins are toxic to cells when they form in or are trafficked to the nucleus-although these early results are being

10

questioned through more systematic assessments of nuclear trafficking and inclusion formation.

1.2.2 Polyglutamine in non-pathogenic contexts

Although our primary interest in polyglutamine stems from its disease-relevance, this thesis would be incomplete without mentioning some other contexts in which polyglutamine or at least high glutamine content is found. Additionally, nature has selected these sequences for a reason, so the role of polyglutamine in these contexts is discussed.

Many proteins in cereal grains (seeds of wheat, barley, oat, and rye) contain high amounts of glutamine (*100-102*). These glutamine rich, proteins are known as storage proteins, because their purpose is to store nutrients for the germinating seedling. In particular, glutamine residues in these proteins act as a source of nitrogen (*103*) which is a crucial nutrient for all plants. (*104*) These storage proteins are accumulated in "protein bodies" which are aggregates of the proteins and are usually sequestered from the cell in a membrane surrounded vesicle. (*105*) In this regard, these protein bodies are not dissimilar from many aggregates in human disease.(*106, 107*) Interestingly, the aggregates seem to provide an answer to a dilemma for seeds: how to store proteins for sustained release of nutrients. The proteins stored in the protein bodies need to be resistant to degradation on a long time scale (weeks) so as to not be degraded during seed formation but can broken down slowly during germination to provide the growing seedling with the stored nutrients. Storing the proteins in an aggregated form in protein

bodies seems to help accomplish this goal of sustained nutrient release. (*108*) In this context, the ability of polyglutamine-rich sequences to aggregate and resist proteolytic breakdown is crucial for survival, in stark contrast to their deleterious role in human disease.

Although the aggregation of the seed storage proteins and their resistance to proteolytic breakdown are beneficial to the plant, these same properties can lead to serious problems in human patients who have a condition known as Celiac Sprue. (*109-111*) Celiac Sprue is a relatively common digestive disorder which is caused by an inflammatory response to glutamine-rich seed storage proteins in genetically susceptible individuals. (*112*) It has been shown that these proteins are resistant to complete degradation by the complement of enzymes in the human gut (*113-115*) and this leads to pathogenicity in Celiac Sprue patients by leaving immunogenic fragments behind after proteolysis. (*114, 115*) Like the proteins in polyglutamine expansion diseases, these proteolytic fragments of the seed storage proteins are rich in glutamine and are even more pathogenic than the full length proteins. (*114, 115*)

Although consumption of the glutamine-rich seed storage proteins can be harmful for Celiac Sprue patients, humans have been consuming these proteins by combining them with water to make dough for over 30,000 years. (*116*) Only recently have we realized that the extensive interprotein hydrogen-bonded network due to glutamines gives rise to many of the properties in dough which make it useful for culinary applications. (*117*)

Although dough is important historically, one could make the argument that we could easily live without it. However, we, and almost every other organism on Earth would not exist if not for another very important class of proteins which are enriched in glutamine or at least have domains enriched in glutamine: transcription factors (TFs). (*118-124*) Transcription factors control the conversion of genetic information contained in the DNA to a signal (mRNA) telling the protein-making machinery what proteins to make. (*123, 124*) All cells contain the same genetic information, and it is thought that the combinatorial use of TFs gives rise to cell-specific functions and cell-differentiation during development. (*125*) The role of glutamine-rich regions (QRRs) in TFs remains unclear.

One hypothesis for the role of QRRs in TFs is that polyglutamine confers conformational flexibility needed for the transcription factors to interact with their binding partners. (*122*) It is known that QRRs are important in protein-protein interactions necessary for TF function. (*120*) Moreover, there is an optimal polyglutamine length for TF function, and it is possible to induce TF activation by adding polyglutamine. (*119*)

Another hypothesis is that the QRRs can serve as a tether between two functionally distinct domains in multi-domain TFs. (*126*) The QRRs hold domains in close proximity while maintaining the conformational flexibility for these domains to orient themselves to perform their functions.

Although the QRRs in TFs serve some important purpose, a likely unintended consequence is that they can also lead to aggregation of these proteins. Spinocerebellar Ataxia 17 is one of the polyglutamine expansion diseases thought to be caused by aggregation of abnormally expanded TATA box binding protein (a transcription factor). (*53*) In addition, many normal length TFs are found in intranuclear inclusions of polyglutamine disease patients. (*80*) This has led to speculation that toxicity comes from transcriptional dysregulation which may arise from sequestration of TFs in neuronal intranuclear inclusions. (*127, 128*) That polyglutamine domains are found in a variety of different protein contexts suggest that these domains have unique functional properties. A characterization of the physicochemical properties of polyglutamine is necessary for understanding both its normal roles and possibly abnormal roles in biological systems.

1.3 Prior Studies on the Biophysical Characterization of Polyglutamine

1.3.1 Establishing polyglutamine's solubility characteristics

Histopathology of polyglutamine disease patients' brains shows the presence of insoluble aggregates of polyglutamine containing proteins. However, this is certainly not the first observation that high glutamine content leads to insolubility. In fact, it has been known since the early 1900's that proteins such as the wheat protein gluten contained percentages of glutamine as high as 37% and were remarkably insoluble in water (*4, 6, 101, 102, 129*) even though glutamine is a polar, and reasonably soluble (*1*) amino acid.

The first biophysical understanding of why peptides rich in glutamine might be insoluble came from the work of Krull and Wall. (*2, 5*) They synthesized random polymers of varying percentages of glutamine and glutamic acid and then examined the solubility of these constructs in aqueous solution as a function of pH. The criterion for solubility was the onset of turbidity: if a 0.2% polymer solution contained in a 10 cm polarimeter tube was visually clear, the solution was deemed soluble. They also performed optical rotary dispersion measurements on these solutions to determine α-helical content. They made several interesting observations: 1) the solubility increased with increasing pH for constructs with less than ~60% glutamine (constructs with 60% glutamine could not be dissolved at all); 2) the solubility of the constructs at neutral pH decreased with increasing glutamine content; 3) the α-helical content decreased with increasing glutamine content; and 4) the apparent pK_a of the carboxyl groups on the

glutamic acids increased with increasing glutamine content (which is not expected for a random coil polymer) but returned to the expected value upon addition of 2 M urea. The authors concluded that charge repulsions due to ionization of the glutamic acids can overcome intermolecular forces which promote aggregation. One of these forces was suspected to be the hydrogen bonding due to the side chain amide of glutamine. The authors also point out that glutamine residues promote structural disorder. Finally, the authors suggest that the pK_a shifts of the carboxylic acid group in polymers of high glutamine content are likely due to a protection of the sidechain from solvent. In Chapter 2, we will show that this is likely because these polymers form collapsed globules in aqueous solution.

1.3.2 Characterizing the structural characteristics of polyglutamine aggregates

After the discovery that uninterrupted glutamine tracts were implicated in neurodegenerative diseases (*60, 65, 71, 74*), interest in the biophysical characterization of polyglutamine increased. One significant technical problem with the study of polyglutamine is that it is insoluble in aqueous solution. For this reason, researchers have used a variety of sequence modifications to improve solubility, under the assumption that these modifications do not alter the underlying properties of polyglutamine. The typical modification is the addition of charged amino acids at the termini.

Perutz and coworkers synthesized constructs of the form $D_2Q_{15}K_2$ to study the structure of polyglutamine in aqueous solution. (*130*) The terminal residues were added to improve solubility at pH 3 while maintaining net neutrality at pH 7. The authors performed circular dichroism (CD) measurements of the peptide in aqueous solution at pH 3 and at pH 7. The CD spectra of polyglutamine solutions at both pH values indicated the peptides were rich in β-sheet, but only the solution at pH 7 showed slow precipitation of insoluble aggregates. They investigated the higher order structure of these precipitated aggregates with electron microscopy, which showed linear, worm-like aggregates of 0.7-12 nm in thickness and varying length. An x-ray diffraction photograph of the same precipitated aggregates suggested a superstructure dominated by cross β-sheet which is oriented parallel to the axis of the linear aggregates. The authors suggested that the forces driving polyglutamine to aggregate are hydrogen bonds between the main-chain amides and the backbone amides, an effect termed polar zipper. (*130, 131*)

Seven years after initially hypothesizing that polyglutamine aggregates form cross-β sheets, Perutz and coworkers revisited their original diffraction data and reinterpreted it to suggest that polyglutamine aggregates form cylindrical β sheets wrapped around a core of water, a structure the authors refer to as water-filled nanotubes.(*132*) In this model, the glutamine sidechains alternate between sticking into the core of the nanotubes, and sticking out into solvent.

Sikorski and Atkins reinterpreted the same X-ray diffraction data. (*133*) They suggested that the polyglutamine aggregates are better characterized as cross-β crystallites where the aggregates are held together by side-chain amide to side-chain amide hydrogen bonds in addition to the traditional β-sheet backbone amide to backbone amide hydrogen bonds. The glutamine sidechains are highly interdigitated leading to very dense aggregates (~1.54 g/cm^3), which is consistent with a crystal structure obtained for the peptide GNNQQNY. (*134*) Sikorski and Atkins maintain that this dense structure is more likely to be proteolytically resistant and is more consistent with insoluble aggregates than the water-filled nanotubes. Sharma et al. performed x-ray fiber diffraction experiments on aggregates of polyglutamine constructs with the form: Q_8, $D_2Q_{15}K_2$, $K_2Q_{28}K_2$, and $K_2Q_{45}K_2$. Their data were also consistent with the aggregates polyglutamine adopting cross-β crystallite geometries. (*135*)

Aggregated polyglutamine was fairly convincingly characterized structurally, but questions remained regarding its structure prior to aggregation. The CD spectra of the constructs used by Perutz and coworkers, $D_2Q_{15}K_2$, showed β-sheet structure in aqueous solution at both pH 3 and pH 7. (*130*) However, only the pH 7 solution showed eventual precipitation of aggregated peptides. One might hypothesize from these findings that monomeric polyglutamine (pH 3) and aggregated polyglutamine (pH 7) are both dominated by β-sheet. This would be an incorrect interpretation because Wetzel and coworkers later showed that the pH 3 solution of these same peptides contained aggregates even though these aggregates did not appear to precipitate on a very long time scale.(*136*) This simply means that the aggregates in the pH 3 solution appear to be

soluble, whereas the aggregates in the pH 7 solution are insoluble, but both forms of aggregates contain high amounts of β-sheet.

1.3.3 Clarifying the terminology of protein aggregation

Before proceeding, it is worth making a few points regarding semantics. In the literature, one sees implicit terminological conflation between aggregation, precipitation, and phase separation. Indeed, the latter term is seldom used, and precipitation and aggregation are generally thought to be synonymous. This conflation originates, at least partially, in the formal structure of microstate partitioning models (MPMs), which will be discussed in Section 1.3.3. Another cause of this conflation is the way in which aggregation is followed or assessed. This was evident in the aforementioned case of Perutz and coworkers where the peptide solution at pH 3 was visually clear and showed no visually observable change with time, but the solution at pH 7 showed obvious precipitation of the sample. (*130*) If the solutions had been interrogated using some sort of sizing assay such as: light scattering, size-exclusion chromatography, or gel electrophoresis; the authors would have likely concluded that both solutions contained aggregates.

In this thesis, when we use the term aggregation we will be referring to a process where a number of molecules associate to form higher order species called aggregates. We assume that the aggregation process leads to kinetically stable (meaning they are long lived on an arbitrarily defined observation timescale) if not thermodynamically stable

aggregates. These aggregates may or may not be soluble. If the aggregates are insoluble, they are precipitates and the process by which these insoluble aggregates form and settle out of solution is known as precipitation. Therefore, precipitates are aggregates but have the additional distinction of being insoluble. The relationship between aggregation and phase separation is not as simple to define, but this will be discussed more in Section 1.4 and Chapter 3.

1.3.4 Characterizing the structural characteristics of monomeric polyglutamine

Wetzel and coworkers devised a protocol for disaggregating polyglutamine samples, which provided confidence that the samples being studied were indeed monomeric.(*136*) In their protocol, polyglutamine is studied in constructs of the form $K_2Q_NK_2$, again under the assumption that the added lysines improve solubility without altering the properties of polyglutamine itself. Instead of dissolving the dry peptide directly in water, the peptide is first dissolved in a 1:1 mixture of trifluoroacetic acid (TFA) and hexafluoroisopropanol (HFIP) which are thought to break all of the inter- and intramolecular hydrogen bonds, effectively dissolving all preformed aggregates. This solution is then dried to a peptide film in a glass container. The peptide is then resuspended initially in water (adjusted to pH 3 with TFA) and the solution is then adjusted to the appropriate buffer condition. Finally, this solution is centrifuged at >50,000×G for at least 3 hours to remove any traces of remaining aggregates and the top 2/3 of the supernatant is removed and used for experiment immediately. CD measurements of polyglutamine samples prepared this way were characteristic of random

coil, meaning that these peptides lacked any well-defined secondary structures. (*137*) This was true regardless of the polyglutamine length. Several other experimental studies corroborated the finding that monomeric polyglutamine is disordered. (*138, 139*) Additionally, simulation results from the Pappu lab also suggested that not only do homopolymeric polyglutamine (i.e. without any charged residues) have no secondary structure preference, but that these molecules preferred collapsed geometries. (*140-142*) Polyglutamine's preference for collapsed structures will be discussed in greater detail in Chapter 2.

When the rigorously disaggregated polyglutamine samples used by Wetzel and colleagues were incubated in solution, the peptides precipitated and the secondary structure, as determined by CD, suggested that these precipitates were enriched in β-sheet. (*137, 143*) EM images showed that these β-sheet rich precipitates are consistent with the structures seen by Perutz and co-workers (*130*) which are generally referred to as amyloid fibrils (*7*). Amyloid is currently[1] defined as protein-based, fibrillar structure showing a cross-β sheet morphology. (*7*) Amyloid is typically identified by binding of the dyes Thioflavin T or Congo red, the latter of which shows a green birefringence upon binding. (*7, 144-147*) The dye-based identification of amyloid must be used with caution, however, as these dyes have been shown to be non-specific in that they are markers of fibrillar structures that need not have β-sheet character. (*148, 149*)

---

[1] The name amyloid derives from the amylase-like dye-binding characteristics of protein fibrils. (7)

In vitro studies have shown that the overall rate of formation of "ordered" precipitates increases with polyglutamine length. (*150, 151*) Finally, the concentration of polyglutamine remaining in solution at the end of an aggregation reaction decreases with increasing polyglutamine length. (*152-154*) The evidence strongly suggests that polyglutamine exists as a disordered monomer in isolation, and undergoes a collective "folding" reaction upon aggregation. Moreover, the increased polyglutamine length leads to a higher thermodynamic and kinetic drive to aggregate.

1.3.5 Thermodynamic and kinetic descriptions of protein aggregation using microstate partitioning models (MPMs) [2]

Prior to discussing attempts to characterize the mechanism of polyglutamine aggregation, it is necessary to provide a description of the theoretical framework generally used. In this MPM framework, which is applied to many protein aggregation systems, one has to define all possible microstates of an aggregation reaction.

Aggregation of proteins typically results in the formation of two types of higher order structures: amorphous aggregates and ordered aggregates such as amyloid fibrils. (*155, 156*) Historically, models for protein aggregation have focused on the formation of amyloid fibrils and these models have largely discounted the formation of amorphous aggregates. The two most likely reasons are that these structures were originally implicated as a possible cause for the disease and that the formation of these structures is

---

[2] This section is adapted from the book chapter: S.L.Crick, R.V. Pappu. (2010).Thermodynamic and Kinetic Models for Aggregation of Intrinsically Disordered Proteins. *Peptide Folding, Misfolding, and Unfolding*. Ed. Reinhard Schweitzer-Stenner. John Wiley & Sons, Hoboken, NJ. In press.

easily followed with simple techniques such as fluorescence enhancement of thioflavin T. (*157-160*) MPMs for protein aggregation are built on a wealth of literature describing the formation of fibrillar structures such as F-actin and sickle cell hemoglobin. (*161-165*) Our discussion of MPMs is not meant to be a comprehensive review of these models because such coverage is already available in the literature. (*32, 166-173*) We will instead provide a critical overview of basic mechanisms and evaluate MPMs in terms of their predictive power and ability to reproduce known characteristics of protein aggregation.

1.3.5.1 MPMs for thermodynamics of linear aggregation

We start with the most intuitive mechanism for the formation of a linear aggregate, which involves growth via monomer addition. This mechanism is shown in Scheme 1. With knowledge of the rate constants of each step and the starting concentration, this simple mechanism can describe an aggregation mechanism with no off-pathway intermediates (where off-pathway means that these species are non-productive towards the formation of the final product). It is impractical to enumerate every elementary step required to form a high molecular weight species like a fibril. Even if such an enumeration were practical, one could never measure the concentration of each species along the pathway, which is a requirement for determining the associated rate constants for each step. As a result, researchers are forced to make simplifying assumptions to this model.

Scheme 1

$$M_1 + M_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} M_2; \quad K_2 = \frac{\left[M_2\right]}{\left[M_1\right]\left[M_1\right]}$$

$$M_2 + M_1 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} M_3; \quad K_3 = \frac{\left[M_3\right]}{\left[M_2\right]\left[M_1\right]}$$

$$\vdots$$

$$M_i + M_1 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} M_{i+1}; \quad K_{i+1} = \frac{\left[M_{i+1}\right]}{\left[M_{i+1}\right]\left[M_1\right]}$$
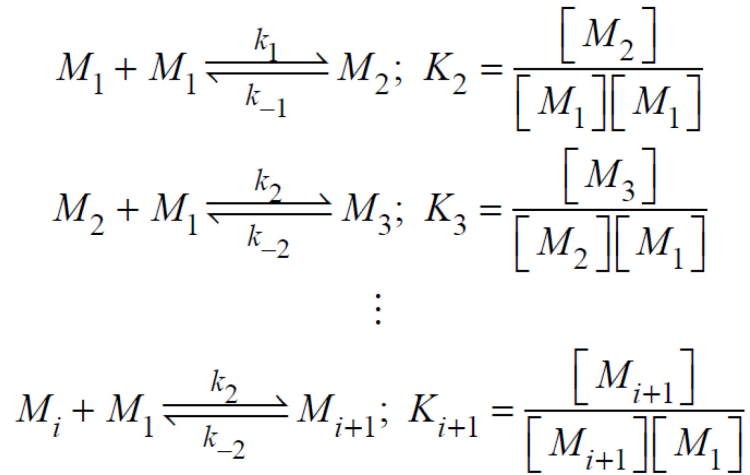
Isodesmic aggregation: One simplification is to assume that every step in the reaction has the same equilibrium constant (K). (*174, 175*) This is known as isodesmic aggregation and the consequences of this simplification are shown in "Equations for isodesmic aggregation". Using this assumption, it can be shown that the concentration of the monomeric species M_1, and therefore, any species along the reaction coordinate is determined simply by the initial monomer concentration C_0 and the equilibrium constant K. Assuming one can measure this monomer concentration M_1 at equilibrium, it is simple to determine K from a concentration dependence of M_1 as a function of C_0. Figure 1.2.A shows a plot of *DA*, the degree of aggregation, as a function of the dimensionless parameter KC_0. This reveals that extensive aggregation cannot be realized unless KC_0 >>1 requiring that for typical concentrations (C_0 << 1 M), K must be large. This is illustrated in Figure 1.2.B.
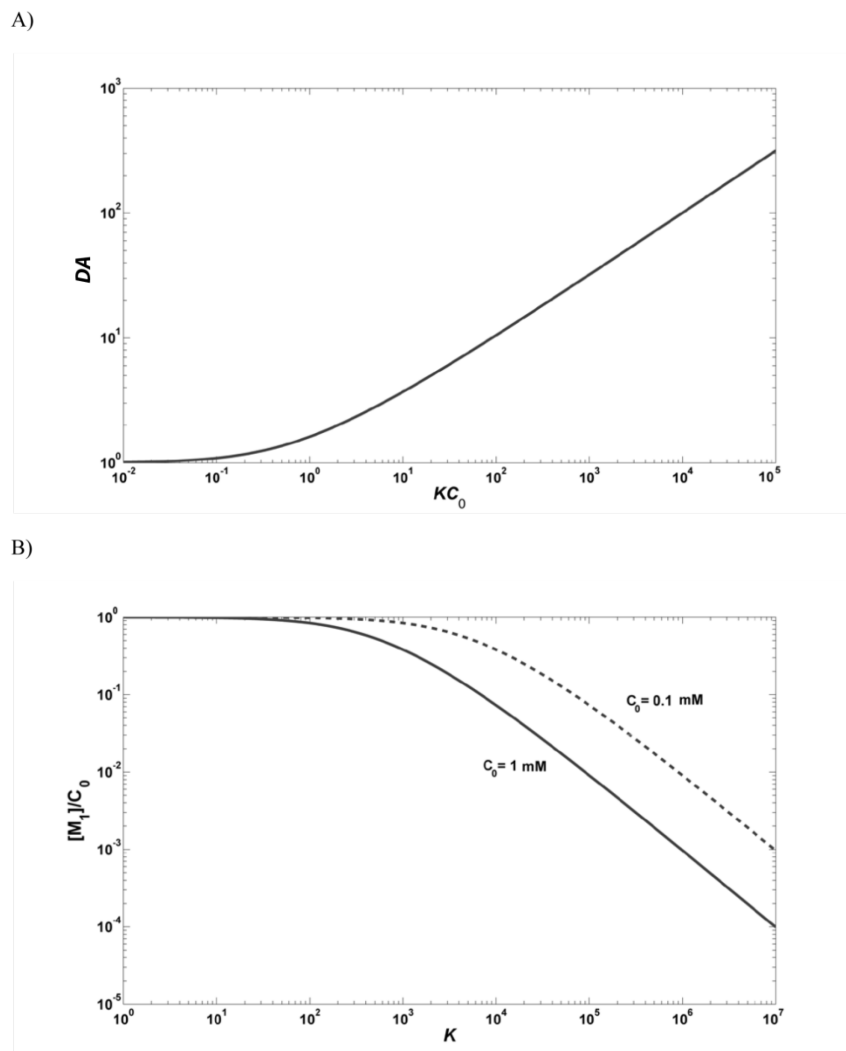
Figure 1.2 A) The degree of aggregation (*DA*) as a function of the initial monomer concentration (C_0) and the equilibrium constant (K) for an isodesmic aggregation mechanism. **B)** The concentration of monomer [M_1] relative to the initial monomer concentration (C_0) remaining in solution at equilibrium in an isodesmic aggregation mechanism as a function of the equilibrium constant (K). Two different initial monomer concentrations are shown: 1 mM (solid line) and 0.1 mM (dashed line). The position of the 1 mM curve relative to the 0.1 mM curve is a reflection of the increased monomer

incorporation into polymer at a given K. This can also be seen from Figure 1.2.A where, for a given K, the higher C_0 will lead to a higher degree of aggregation.

Equations for Isodesmic Aggregation

$$K_i = K \text{ for } 2 \leq i < \infty$$

$$[M_2] = K[M_1]^2$$

$$[M_3] = K[M_2][M_1] = K^2[M_1]^3$$

$$\vdots$$

$$[M_{i+1}] = K^i[M_1]^{i+1}$$

$$\text{For } K[M_1] < 1$$

$$C_o = \sum_{i=1}^{\infty} i[M_i] = \frac{[M_1]}{\left(1 - K[M_1]^2\right)} \text{ is the initial concentration}$$

$$\text{Solving for } [M_1] \text{ yields: } [M_1] = \frac{\left(2K + \frac{1}{C_o}\right) - \sqrt{\left(2K + \frac{1}{C_o}\right)^2 - 4K^2}}{2K^2}$$

$$C_P = \sum_{i=1}^{\infty} [M_i] = \frac{[M_1]}{\left(1 - K[M_1]\right)^2} \text{ is the partition function}$$

$$\langle i \rangle = DA = \frac{C_o}{C_P} = \frac{1}{1 - K[M_1]}$$

Isodesmic aggregation is the simplest possible mechanism and that remains its strength. However, it fails to capture several key features of the presumed linear aggregation of many proteins. The weaknesses are multi-fold: The concept of a

saturation concentration[3] (C_{sat}) is not defined nor is the possibility of seeding a reaction using pre-formed aggregates. As can be seen in Figure 1.2A, the concentration of higher-order aggregates will increase with the initial monomer concentration C_0 for a given equilibrium constant K. Addition of preformed aggregates to a reaction containing a monomer concentration above the saturation concentration will cause growth of the aggregates and a decrease in the monomer concentration. (*176-179*) In isodesmic aggregation, the monomers of the pre-formed aggregates would repartition into other species and a new equilibrium would be established whereby the concentration of monomer would actually increase.

Linked isodesmic processes: To explain the cooperativity observed in actin aggregation, Oosawa and coworkers considered two isodesmic processes that are linked such that the first process models the consequences of a free energy profile as a function of aggregate size for which $G(n) - G(n–1) > 0$ for $n < n*$ and a second process for which, $G(n) - G(n–1) < 0$ for $n > n*$. (*161, 162, 173*) This species of size $n*$ is the nucleus and represents a peak on the free energy profile of G versus n. Oosawa's mechanism, depicted in Scheme 3, is referred to as homogeneous nucleation or the nucleation-elongation mechanism.

---

[3] In protein aggregation (and any phase separation process), there exists a saturation concentration (C_{sat}) below which phase separation will not occur spontaneously. Assuming phase separation has occurred and has reached equilibrium, the saturation concentration is also the amount of soluble material that will remain dissolved in solution. This concept will be rigorously addressed in Section 1.4.

Scheme 3

$$M_1 + M_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} M_2; \quad K_2 = \frac{[M_2]}{[M_1][M_1]}; \quad [M_2] = K[M_1]^2$$

$$M_2 + M_1 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} M_3; \quad K_3 = \frac{[M_3]}{[M_2][M_1]}; \quad [M_3] = K^2[M_1]^3$$

$$\vdots$$

$$M_{n*-1} + M_1 \underset{k_{-n*}}{\overset{k_{n*}}{\rightleftharpoons}} M_{n*}; \quad K_c = \frac{[M_{n*}]}{[M_{n*-1}][M_1]}; \quad [M_{n*}] = K_c K^{n*-2}[M_1]^{n*}$$

$$\vdots$$

$$M_i + M_1 \underset{k_{-i}}{\overset{k_i}{\rightleftharpoons}} M_{i+1}; \quad K_{i+1} = \frac{[M_{i+1}]}{[M_i][M_1]}; \quad [M_{i+1}] = (K_c)^{i+1-n*} K^{n*-2}[M_1]^i \text{ for } i \geq n*$$

$$\text{or } [M_{i+1}] = \sigma(K_c)^{i-1}[M_1]^i \text{ for } i \geq n*$$

$$\text{where } \sigma = \left(\frac{K}{K_c}\right)^{n*-2}$$
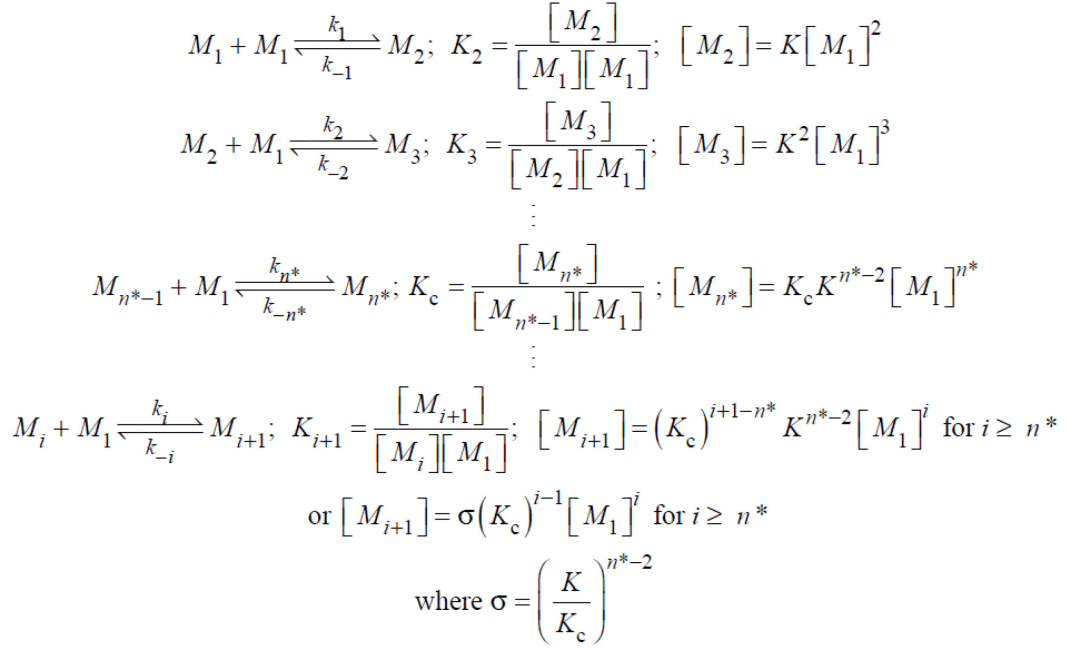
By introducing a set of unfavorable steps (followed by favorable steps) into the reaction described in "Equations for isodesmic aggregation", one models two linked isodesmic processes through the cooperativity parameter σ. Here, K is less than unity and K_c is greater than unity and as a result σ is < 1. We can obtain the expression for the total monomer concentration to demonstrate that this leads to the appearance of a saturation concentration C_{sat}. Since $C_o = [M_1] + \sum_{i=2}^{\infty} i\sigma K_c^{-1}(K_c[M_1])^i$, which for $\sigma \ll 1$ and $K <$

K_c yields $C_o \approx [M_1] + \dfrac{\sigma[M_1]}{(1 - K_c[M_1])^2}$. That the Oosawa model supports the existence of a

saturation concentration is demonstrated in Figure 1.3A. For the purpose of illustration, we set $n* = 4$ and $K_c = 10^5 \, \text{M}^{-1}$ with $K = 1 \, \text{M}^{-1}$. If $C_o < K_c^{-1}$, then the model yields nearly a one-to-one correspondence between the monomer pool ($[M_1]$) and C_o. For $C_o > K_c^{-1}$, the

monomer concentration shows a plateau value and this becomes the fraction of C_o that does not get incorporated into higher molecular weight species. In the Oosawa model K_c^{-1} is the saturation concentration because it fits both criteria outlined on page 1 for the definition of C_{sat}. No higher order aggregates form until the C_{sat} is crossed. This is clear from a plot of the degree of aggregation $DA = \dfrac{1}{1-K_c[M_1]}$ as a function of the initial monomer concentration, shown in Figure 1.3B.

A)
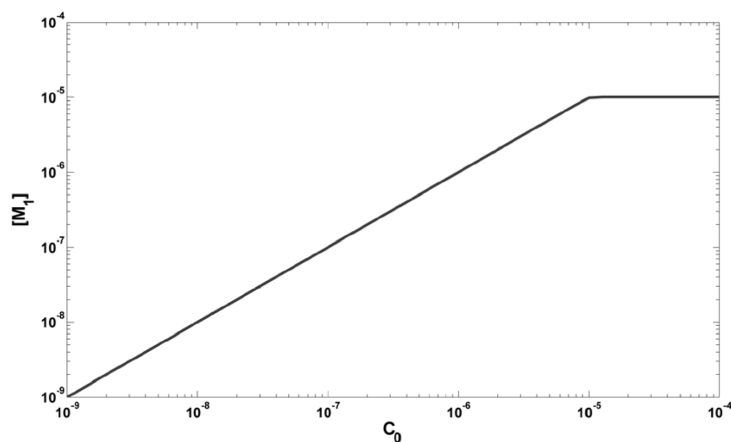
B)
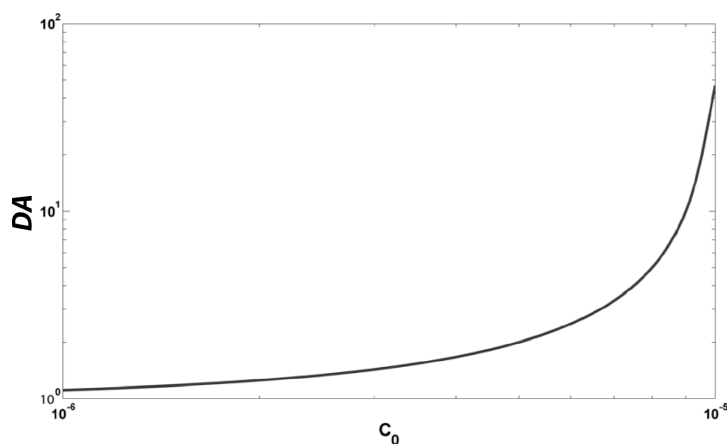
Figure 1.3 A) The concentration, in molar units, of monomer $[M_1]$ remaining in solution at equilibrium in a nucleation-elongation aggregation mechanism with a nucleus size of 4, a pre-nucleation equilibrium constant (K) of 1 M^{-1}, and a post-nucleation equilibrium constant (K_c) of 10^5 M^{-1}. This leads to a predicted saturation concentration of 10^{-5} M, which is evident by the plateau in $[M_1]$ beginning at the saturation concentration. B) The degree of aggregation as a function of the initial monomer concentration (C_o) for the aggregation mechanism and conditions described in Figure 1.3.A. It is expected that the degree of aggregation will be low at values of C_o less than the saturation concentration, and will start to increase dramatically at concentrations near the saturation concentration

and higher. This is due to the fact that, because the concentration of $[M_1]$ remains constant at equilibrium at C_o higher than the saturation concentration, all additional monomer must be incorporated into the high order polymers as C_o is increased.

Oosawa's model provided one of the original descriptions of the thermodynamics of aggregation (polymerization in his parlance) in the MPM framework. This framework can be generalized to include additional steps such as condensation of fibrils, fragmentation of higher molecular weight species, and the effects of conformational changes. (*166, 180*) Roberts and coworkers have included the effects of conformational changes through the so-called Lumry-Eyring nucleated polymerization (LENP) model. (*181, 182*) All of these nuances require the inclusion of additional microstates and steps to describe the conversion between these states. At a minimum, each additional microstate requires three extra parameters namely two rate constants and the activity (concentration) of the new species. This poses severe challenges because one cannot use standard data sets to determine the requisite parameters.

1.3.5.2 MPMs for kinetics of linear aggregation via homogenous nucleation and elongation

In the previous section we considered the equilibrium behavior of nucleation-elongation processes. Although this is useful for illustrating the concentration dependence of aggregation, it ignores the rich, time-dependent evolution of nucleation-elongation mechanisms.

One feature of nucleation-elongation processes is the presence of a lag phase associated with the formation of species larger than the nucleus. A second feature is shortening of the lag-phase by seeding the reaction with pre-formed aggregates. (*163, 164, 168*) It has been observed that many protein aggregation mechanisms have both of these features, which makes the nucleation-elongation mechanism particularly appealing for modeling these processes. (*179, 183, 184*)

To simplify the modeling of the time dependence of a nucleation-elongation process, two assumptions are generally made. These are: 1) the monomer (*M*) is in rapid equilibrium with the nucleus (*N*) with equilibrium constant (K_{PN}), and 2) the addition of the monomer to the nucleus or any species larger than the nucleus leads to the formation of a fibril and is irreversible. (*169, 170, 172, 173*) Applying these simplifying assumptions yields the pre-equilibrated nucleus model shown in Scheme 4.

Scheme 4

$$n * M \xrightleftharpoons{K_{PN}} N$$
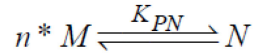
this is the pre-equilibration step for forming the nucleus

$$M + N \xrightarrow{k_{1PN}} F$$
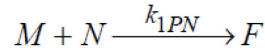
this step refers to irreversible elongation of the nucleus

$$M + F \xrightarrow{k_{2PN}} F$$
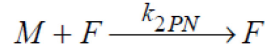
irreversible loss of monomer to the fibril

$$[N] = K_{PN} [M]^{n*}$$

concentration of the nucleus in the pre-equilibrium model

$$\frac{d[M]}{dt} = -k_{1PN} K_{PN} [M]^{n*+1} - k_{2PN} [M][F]$$

rate of loss of monomer

$$\frac{d[F]}{dt} = k_{1PN} K_{PN} [M]^{n*+1}$$

rate of growth of fibril

In this mechanism, $n*$ is the number of monomers in the nucleus, $[M]$ denotes the concentration of monomers, N is the nucleus, and F is the fibril. These equations are co-dependent on the concentration of the monomer and consequently, monomer loss kinetics alone cannot be used to determine both the pre-equilibrium constant (K_{PN}) and the rate constant for elongation of the nucleus, k_{1PN}. Instead, one must follow both monomer loss and the formation of fibril. These equations can be numerically integrated with $[M] = [M]_0$ and $[F] = 0$ at time $t = 0$ for an unseeded reaction.

The monomer-loss and fibril growth kinetics using a pre-equilibrated nucleus model with a monomeric nucleus are shown in Figures 1.4A and 1.4B, respectively. The

equilibrium and rate constants used were those determined by Wetzel and co-workers for the aggregation of polyglutamine ($K_{PN} = 2.6 \times 10^{-9}$ and $k_{1PN} = k_{2PN} = 11,400 \ s^{-1}M^{-1}$). (*150, 151*) As expected for a nucleation-elongation process, both the monomer-loss and fibril growth kinetics show a pronounced lag phase before aggregation is readily apparent.

A)
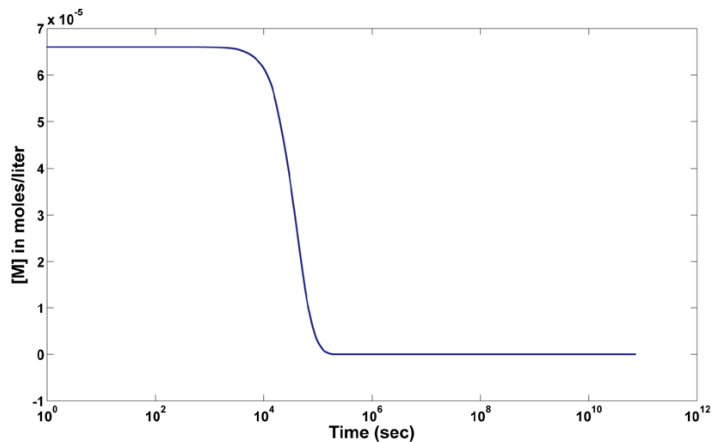
B)
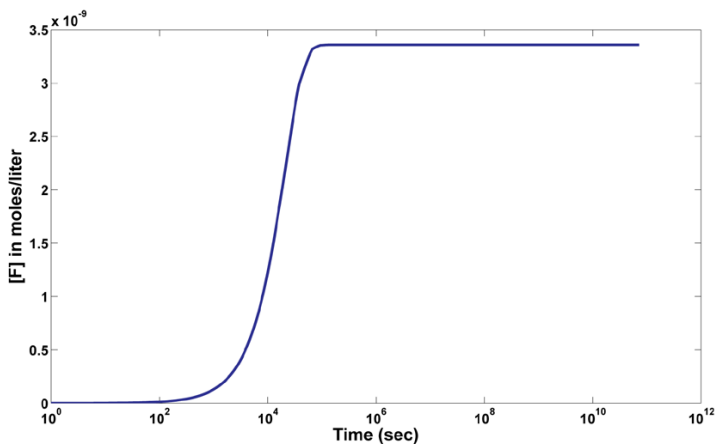
Figure 1.4 A) The monomer concentration [*M*] as a function of time in a pre-equilibrated nucleus model for nucleation-elongation. The pre-equilibrium constant (K_{PN}), rate constants, and nucleus size (*n**) used were those determined by Wetzel and co-workers for the aggregation of polyglutamine ($K_{PN} = 2.6 \times 10^{-9}$, $k_{1PN} = k_{2PN} = 11,400 \ s^{-1}M^{-1}$, $n* = 1$). The starting concentration was 66 μM. There is a pronounced lag-phase before the

monomer concentration begins to decrease. **B)** The fibril concentration [F] as a function of time in the same model and conditions described in Figure 1.4A. As with the monomer concentration in Figure 1.4A, there is a pronounced lag-phase before the fibril concentration begins to increase. The fibril growth is concomitant with monomer loss.

For a given set of equilibrium constants and rate constants, the rate of monomer loss in the pre-equilibrated nucleus model will decrease with increasing nucleus size and increase with increasing monomer concentration. This is illustrated in Figures 1.5A and 1.5B. Also, the initial rate of fibril formation depends on the nucleus size. This is illustrated in Figure 1.5C. Another feature of a nucleation-elongation reaction is that the lag time can be shortened with increasing amounts of preformed aggregate. This effect is illustrated in Figure 1.5D. As the starting percentage of preformed aggregate increases, the lag time is shortened dramatically.

A)
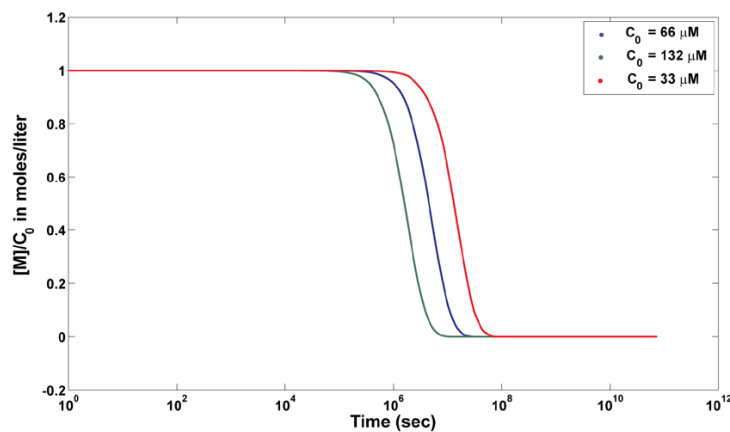
B)
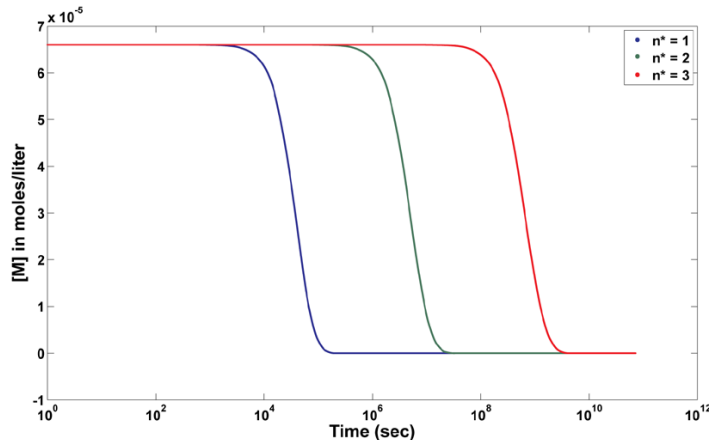
Figure 1.5 A) The relative monomer concentration $[M]/C_0$ as a function of time in a pre-equilibrated nucleus model for nucleation-elongation shows a dependence on the starting monomer concentration. In this figure, $K_{PN} = 2.6\times10^{-9}$ M^{-1}, $k_{1PN} = k_{2PN} = 11,400$ s^{-1}M^{-1}, $n^* = 2$. The lag phase decreases with increasing C_o. **B)** An investigation of the effect of the nucleus size on the monomer concentration as a function of time for the conditions shown in Figure 1.4A with $C_o = 66$ μM. An increasing nucleus size greatly increases the

lag phase. For $n^*=$ 1, 2 and 3, K_{PN} was set to be K_{PN}=2.6×10^{-9}, 2.6×10^{-9} M^{-1}, 2.6×10^{-9} M^{-2}, respectively.

C)
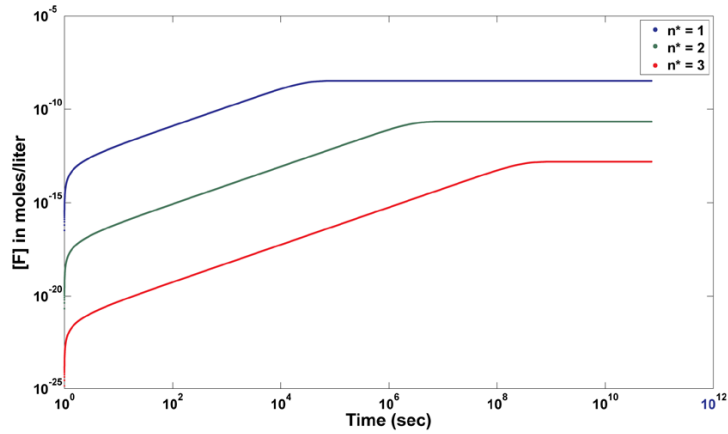
D)
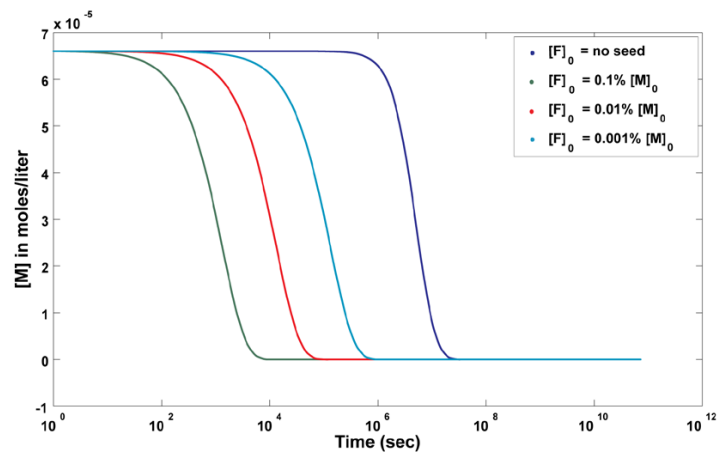
Figure 1.5 C) An investigation of the effect of the nucleus size on the fibril concentration as a function of time for the conditions shown in Figure 1.5.B. The rate of fibril growth is fastest at early times. The concentration of fibril remaining at the end of the polymerization is dependent on the nucleus size, with smaller nuclei leading to higher concentrations of fibrils. **D)** The effect of adding various concentration of pre-formed

fibril [F] on the monomer concentration as a function of time for the conditions shown in Figure 1.4.A with C_o = 66 μM. The addition of pre-formed fibril to the beginning of a polymerization reaction reduces the lag phase. In the pre-equilibrated nucleus model, it is not possible to completely eliminate the lag phase with seeding.

Going beyond pre-equilibration models: There are a number of other mechanisms proposed for fitting kinetic data that give similar kinetic behavior to the nucleation-elongation mechanism. Many of these have been described and evaluated by Bernacki and Murphy. (*169*) Even though these mechanisms are differently formulated, they all share the feature that there is a rate-limiting step prior to the formation of the fibrillar species.

1.3.6 Application of MPMs to polyglutamine systems

Wetzel and coworkers used the pre-equilibrated nucleus variant (Scheme 4 in Section 1.3.3.2) of the Oosawa nucleation-elongation model to characterize the aggregation kinetics of polyglutamine. (*143, 185*) Their goal was to determine the size of smallest species (what they call a critical nucleus) which needs to form before polyglutamine aggregation occurs rapidly. (*143, 185*) Before applying the pre-equilibrated nucleus model described in Scheme 4 in Section 1.3.3.2, Wetzel and coworkers made several simplifying assumptions: 1) if one restricts data analysis to only the initial portion of the kinetic data (they arbitrarily chose to analyze the first 10% of the data) then one can assume that the total monomer concentration [M] is equal to the initial

monomer concentration $[M]_0$, and 2) if the majority of monomer loss occurs through addition to growing fibrils and not through nucleation then one can assume that rate of monomer loss is given by: $\frac{d[M]}{dt} \approx -k_{2PN}[M][F]$. Because $[F]$ can be solved for directly from the fibril growth rate equation $\frac{d[F]}{dt} = k_{1PN}K_{PN}[M]^{n^*}$, the monomer loss equation can also be solved analytically to give the concentration of monomer remaining in solution as a function of time given by the equation $[M] = [M]_0 - \frac{k_{2PN}k_{1PN}K_{PN}}{2}[M]_0^{n^*+2}t^2$. In this model, a plot of $[M]$ versus t^2 yields a straight line with a slope equal to $-\frac{k_{2PN}k_{1PN}K_{PN}}{2}[M]_0^{n^*+2}$. If the slopes are determined for multiple data sets starting at different initial concentrations of monomer, then a plot of log(-slope) versus log($[M]_0$) will have a slope equal to n^*+2. Wetzel and coworkers used this methodology to determine the critical nucleus size (n^*) for the constructs $K_2Q_NK_2$ (for $N = 28$, 36, and 47) where N is the number of glutamines.

To measure the monomer loss, Wetzel and coworkers centrifuged aliquots of an aggregation reaction at various time points and then measured the concentration of soluble material remaining in solution using HPLC. They made another assumption that the only material remaining soluble in solution was monomeric. By using the aforementioned protocol for determining the critical nucleus size, Wetzel and co-workers suggested that the critical nucleus size (n^*) is 0.98 for $K_2Q_{28}K_2$, 0.68 for $K_2Q_{36}K_2$, and 0.59 for $K_2Q_{47}K_2$. They interpreted this to mean that the critical nucleus size is 1 and that a highly unfavorable conformation of the polyglutamine monomer acts as a template for

further aggregate growth. Vitalis and Pappu have shown that application of the equilibrated nucleus model leads to fractional values for estimates of n^* when the underlying process diverges from being concordant with homogeneous nucleation and incorporates heterogeneities such as oligomers. (*186*)

To estimate the pre-equilibrium constant for the formation of this nucleus (K_{PN}), Wetzel and co-workers first measured the growth rate constant (k_{2PN}) of pre-formed aggregates of $K_2Q_{47}K_2$ added to a solution of disaggregated monomers. (*185*) They made another assumption that the rate constant for growth of the nucleus (k_{1PN}) was equal to the rate constant for growth of the aggregates. With a value for these rate constants and knowledge of n^*, the authors were able to determine K_{PN} from their model where a plot of [M] versus t^2 yields a straight line with a slope equal to $-\dfrac{k_{2PN}k_{1PN}K_{PN}}{2}[M]_0^{n^*+2}$. They calculated the pre-equilibrium constant for the formation of the critical nucleus for $K_2Q_{47}K_2$ to be 2.6×10^{-9}, indicative of a highly unfavorable process. They then explain the increased aggregation rates as a function of polyglutamine length by suggesting the formation of the nucleus becomes more favorable as the length of the polyglutamine stretch increases, although this was never tested. They also suggest that the unfavorable conformation which a monomer needs to adopt is a β-sheet conformation because this appears to be the structure the monomer adopts in the aggregates. Simulations from the Pappu lab suggested that formation of β-sheet rich structures in homopolymeric polyglutamine is very unfavorable, but that formation of β-sheet does not become more favorable with increasing chain length as predicted by Wetzel and coworkers. (*187*)

Moreover, polyglutamine can spontaneously dimerize without forming β-sheet rich structures (*142, 187*), and, if one restricts the polyglutamine monomers to β-sheet rich conformations, the spontaneity of dimerization decreases. (*187*)

Murphy and coworkers reexamined the aggregation of polyglutamine using the construct $K_2Q_{23}K_2$. (*188*) Using static and dynamic light scattering, size-exclusion chromatography, and a loss of soluble material assay, they found that soluble aggregates are present well before insoluble aggregates are formed. This lies in direct contrast to Wetzel's pre-equilibrated nucleus model where one would only see monomers and higher order, insoluble aggregates. Bernacki and Murphy subsequently showed that a variety of models (including one that is a very simple polynomial which has no mechanistic interpretation) can be used to fit the Wetzel kinetic data, suggesting that determining the mechanism of aggregation by fitting MPMs to kinetic data is a severely underdetermined problem. (*169*)

We propose that MPMs are inherently inferior to simpler thermodynamics descriptions of aggregation based on inferences from phase diagrams. The number of measurements needed to adequately constrain the fitting is overwhelming and generally inaccessible, whereas knowledge of a phase diagram or even regions of the phase diagram helps narrow down the family of conceivable mechanisms. For this reason, we study the process of polyglutamine aggregation from the vantage point provided by polymer physics and solution phase thermodynamics. We will discuss the salient features of these disciplines in the next section.

1.4 Polymer Physics for Protein Aggregation[4]

1.4.1 Introduction

Recent reviews have suggested that the phase-separated state might be the globally stable state for all proteins. (*189-191*) This does not contradict Anfinsen's thermodynamic hypothesis because the issue of stability cannot be decoupled from protein concentration, although this has not precluded the depiction of confusing cartoons of free energy landscapes that disregard the effects of protein concentration. In dilute solutions of foldable proteins, the folded state must minimize the free energy of the protein plus solvent system, thus giving rise to a homogeneously mixed protein solution. For polyglutamine and other proteins that do not have a well-defined folded state, the situation is more complicated. Irrespective of how / why stability is realized by these proteins in dilute solutions, the situation is fundamentally different when protein concentrations increase, because the effects of homotypic intermolecular interactions will eventually lead to the thermodynamically stable, phase-separated state.

The remainder of this section is organized as follows: We start with a summary of a thermodynamic description for phase separation. We then contrast this description with the aforementioned microstate partitioning models that are common in the protein aggregation literature. (*32, 166-172*)

---

[4] This section is adapted mainly from Pappu, R. V., Wang, X., Vitalis, A., and Crick, S. L. (2008) A polymer physics perspective on driving forces and mechanisms for protein aggregation, *Arch Biochem Biophys 469*, 132-141.

1.4.2 Thermodynamics of protein aggregation – the phase diagram approach

All proteins are polymers and protein aggregation can eventually lead to phase separation. Casting aggregation of polyglutamine in terms of a polymer phase separation process allows us to touch base with the rich theoretical framework of polymer physics. (*192-204*) The properties of polymers are determined by the nature of the subunits comprising the polymer chain and the nature of the solvent. For polyglutamine, the sub-unit is the amino acid glutamine and the solvent is generally aqueous. For a given polymer-solvent combination, the conformational and phase behavior of a polymer can be understood from the interplay between chain-chain interactions, chain-solvent interactions, and solvent-solvent interactions. (*192, 201*) Intrinsically flexible polymers transition between globule and coil states as the solvent quality is altered either through changes in thermodynamic parameters such as temperature and hydrostatic pressure or by altering solution conditions through the addition of cosolutes, salts, or changes in pH. (*205, 206*)

In solutions where the chain-chain interactions are favored over chain-solvent interactions, polymers form compact, globular conformations to minimize the polymer-solvent interface and the solvent quality is characterized as poor. (*192, 201*) In solutions where the chain-solvent interactions are preferred to chain-chain interactions, polymers prefer the swollen coil state whereby the polymer-solvent interface is maximized and the solvent quality is characterized as good. (*192, 201*) Polymers in a poor solvent will phase

separate if their concentration is high enough, whereas polymers in a good solvent will form will-mixed solution at all concentrations.

The driving forces that cause a polymer to collapse in a poor solvent are similar to the forces that cause polymers to self-associate and eventually phase separate. In our discussion, we will consider temperature as the dial for adjusting solvent quality. Like most small-molecule solutes, polymers generally, but not always, have a higher solubility at higher temperatures (*i.e.* C_{sat} increases with increasing temperature). These systems are said to have upper-critical solution temperatures (UCST's). However, there are many polymer-solvent systems for which the polymer solubility increases with decreasing temperature. These systems are said to have lower-critical solution temperatures (LCST's). This seemingly counterintuitive behavior likely arises due strong polar interactions (*e.g.* hydrogen bonds) between either the polymer and the solvent or between solvent molecules. (*207-209*) The LCST can also result from compressibility effects (i.e. volume changes upon mixing) of the polymer. (*208-210*) Almost every protein is capable of forming hydrogen bonds, and they are found in a hydrogen-bonding solvent (water), so one might reasonably assume that aqueous protein solutions might exhibit an LCST. In the remaining discussion of this section, we will be discussing polymer solutions with a UCST, but it is important to remember that more exotic phase diagrams (including phase diagrams with both an LCST and a UCST) do exist. For an excellent discussion of the different types of phase diagrams which might exist for polymer solutions see reference (*211*).

A generic phase diagram for a polymer with an upper critical solution temperature (UCST) is shown in Figure 1.6. (*201*) Solvent quality is controlled by temperature (T). As T increases, the solvent quality improves until eventually a temperature is reached above which the solvent is good for all polymer concentrations. T_θ is the crossover or theta temperature and represents the temperature at which the chain-chain interactions exactly balance the chain-solvent interactions in mean-field descriptions of polymer solutions. The abscissa, labeled ϕ, denotes the volume fraction of polymer in a solution. Low values of ϕ represent dilute solutions.
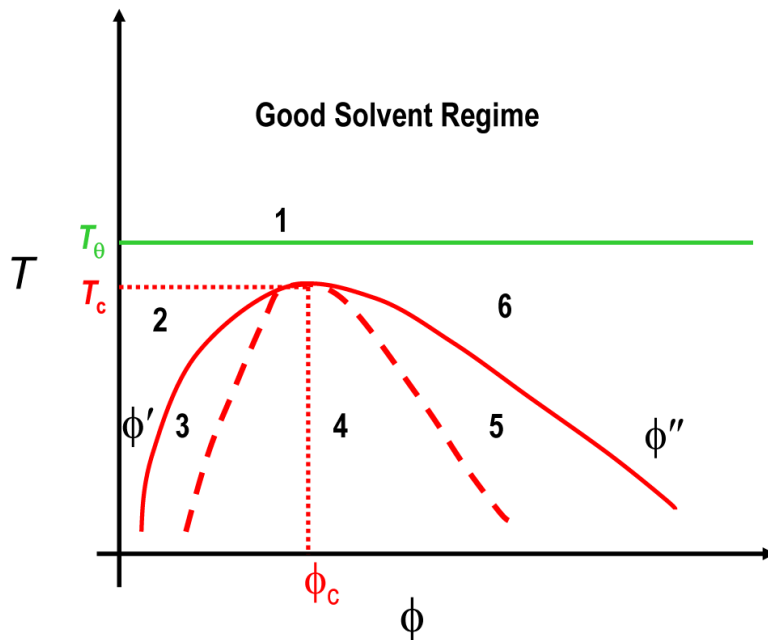
Figure 1.6 Archetypal phase diagram for a polymer solution with an upper critical solution temperature (UCST). The ordinate denotes improving solvent quality expressed as temperature. At the theta-temperature, T_θ, and beyond (good solvent regime) no phase separation is observed. For $T < T_\theta$, a homogeneous mixed phase of polymer in solvent is formed in region 2, and of solvent in polymer in region 6. Conversely, phase separation is

realized in regions 3, 4, and 5. The solid red curve denotes the binodal, while the dashed red curve denotes the spinodal. The point defined by (ϕ_c,T_c) is known as the critical point.

There are six distinct regions in Figure 1.6. Region 1 corresponds to the good solvent regime. In this region, phase separation does not occur because chain-solvent interactions are favored over chain-chain interactions. The solid red curve is called the binodal and is synonymous with coexistence curve for a binary mixture.

Let $\Delta G_{mix}(\phi, T)$ denote the free energy of mixing for a binary mixture of polymers in a low molecular weight solvent. The binodal is an envelope of points representing an equilibrium condition where, for a given temperature, the protein-rich phase at a protein concentration of ϕ'' co-exists with a protein-thin phase at a protein concentration of ϕ'. Mathematically, the binodal is defined as the set of points for which $\frac{d\Delta G_{mix}}{d\phi}(\phi', T) = \frac{d\Delta G_{mix}}{d\phi}(\phi'', T)$. If the combination of T and ϕ places the polymer + solvent system below the binodal, then the system undergoes spontaneous phase separation. The kinetic mechanism of spontaneous phase separation will depend on the precise location of the solution under the binodal and this location is referred to as the quench depth. The dashed curve in Figure 1.6 is the spinodal, which is the envelope of points for which $\frac{d^2\Delta G_{mix}}{d\phi^2}(\phi, T) = 0$. Below the spinodal line, $\frac{d^2\Delta G_{mix}}{d\phi^2}(\phi, T) < 0$ and the homogeneously mixed state is unstable, whereas between the binodal and spinodal $\frac{d^2\Delta G_{mix}}{d\phi^2}(\phi, T) > 0$ and the homogeneous mixed state is metastable. This means that a

solution in the metastable region (under the binodal but above the spinodal) is in a local free-energy minimum, and will require crossing an energy barrier to reach the thermodynamically-favored, phase-separated state. Conversely, a solution in the unstable region (under the spinodal), will phase separate without crossing an energy barrier.

For all points below the binodal, the solution will phase separate into two phases: a solvent rich phase, and a polymer rich phase. (*201*) The concentration of polymer in the solvent rich phase for a given temperature is known as the saturation concentration[5] (ϕ_s). The temperature dependence of this quantity determines the low concentration arm of the binodal (ϕ' in Figure 1.6) and this curve is also called the saturation curve. For a given temperature, ϕ_s can be determined using cloud point measurements wherein the solution will become cloudy at ϕ_s as the temperature decreases below the binodal. (*212*) Another technique for determining ϕ_s is to measure the concentration of polymer remaining in the solvent rich phase after phase separation has occurred at a given temperature. (*213*) We will apply the latter technique to aqueous solutions of polyglutamine in Chapter 3.

In the protein aggregation literature, the saturation concentration is often incorrectly referred to as the critical concentration. (*32, 150, 151, 168*) The critical point (ϕ_c, T_c) on the phase diagram is the point where the binodal and spinodal coincide and is characterized by conformational and concentration fluctuations on all length scales. (*201*)

---

[5] In this section, we use ϕ_s to denote the saturation concentration, as opposed to C_{sat}, to be consistent with the plot in Figure 1.5. The symbol ϕ is reserved for concentrations in units of volume fraction and C_{sat} is reserved for concentrations in units of molarity.

For homopolymers, the chain length determines the location of the binodal and spinodal; for block copolymers and heteropolymers (such as proteins) the ratio of solvophilic to solvophobic groups is an additional parameter that contributes to the location of the binodal and spinodal. (*200*)

Region 2 lies to the left of the binodal. In this sub-saturated region, $\phi < \phi_s$, and the solution is too dilute for phase separation to be thermodynamically favored. Instead, homogeneously mixed globules or mesoglobules[6] characterize the solution. If the concentration is increased beyond the binodal, the solution is in either Region 3 or 4 and is supersaturated. In Region 3 between the binodal and spinodal, the solution is metastable. This is the nucleation-elongation regime where, for a given temperature, the concentration fluctuations must satisfy the condition $(\langle\phi^2\rangle - \langle\phi\rangle^2)^{1/2} > |\Delta\phi|$ where $|\Delta\phi|$ is magnitude of the gap between the binodal and the spinodal. These concentration fluctuations are needed to form the nucleus or nuclei that have the same composition as the new thermodynamically favored phase separated state so this new phase can grow within the old, homogeneously mixed metastable state. Region 4 lies below the spinodal and within this region there is no barrier to phase separation. As a result, phase separation in this region is thermodynamically downhill and is kinetically limited only by the diffusion of individual chains or clusters of chains. (*214*) In Region 5, the solution is concentrated and there exists a barrier for growing the soluble phase inside the polymer precipitate. Finally, Region 6 represents a stable, polymer rich phase with low molecular weight solvent dispersed in it.

---

[6] Mesoglobules are clusters / oligomers of globules and may be viewed as globules of globules

It is worth reiterating that if the combination of ϕ and T for a given solution condition places the polymer solution to the left of the binodal (Region 2 in Figure 1.6), phase separation will not occur. If (ϕ,T) places the solution between the binodal and spinodal (Region 3 in Figure 1.6), phase separation requires the formation of one or more barrier-limited species, referred to as the nucleus (if the process is homogeneous) or nuclei (if the distribution of barriers is heterogeneous). Conversely, if (ϕ,T) places the solution inside the spinodal, even the smallest perturbation in ϕ causes phase separation and the rate of phase separation is limited only by inter- and intramolecular diffusion. This mechanism is referred to as spinodal decomposition.

The preceding discussion is important because it emphasizes the extent of information one can glean from full knowledge of a polymer's phase diagram. This demonstration defies the anecdotal view that thermodynamic descriptions do not provide insights regarding mechanisms of polymer aggregation. To the contrary, knowledge of the phase diagram or even parts of the phase diagram such as the low concentration arm of the binodal and spinodal or the binodal alone (*213*) will narrow the range of mechanisms applicable for the aggregation of specific polymer + solvent systems. Phase diagrams are available for a variety of polymer + solvent combinations (*201*), but, to our knowledge, phase diagrams are unavailable for any of the important disease-related, aggregation-prone proteins in aqueous milieus. The information provided by a phase diagram would be highly useful for characterizing the properties and aggregation of these proteins.

1.5 Summary

This chapter was designed to demonstrate that polyglutamine plays a role in a variety of biological contexts. We are specifically interested in polyglutamine's apparently causative role in neurodegenerative disease pathogenesis. It is clear that polyglutamine content is a major determinant of a protein's solubility in aqueous solutions. The goal of this thesis is to understand: 1) why does increased polyglutamine content lead to a decrease in solubility, and 2) when a polyglutamine-rich protein does phase separate, how might it do so? We seek answers to these questions using the framework of polymer physics. The rest of the thesis is organized as follows.

In Chapter 2, we characterize the solvent quality for monomeric polyglutamine in aqueous solution using fluorescence correlation spectroscopy (FCS) and theories borrowed from polymer physics. This, along with insights from computer simulation, provides a description of the sizes and shapes monomeric polyglutamine might adopt in aqueous solution. In addition, it allows us to understand why polyglutamine phase separates into insoluble aggregates in aqueous solution if the peptide concentration is higher than the saturation concentration.

In Chapter 3, we present a strategy for determining the phase diagram of aqueous polyglutamine solutions, and then we present a temperature-concentration phase diagram for polyglutamine of two different chain lengths in two different sequence contexts. This illustrates how sequence context, specifically the presence of charge, and the

polyglutamine chain length can affect phase behavior and driving forces for aggregation. We conclude Chapter 3 with a discussion of the implications of the phase diagram for the mechanism of phase separation of aqueous polyglutamine solutions.

In Chapter 4, we discuss classical nucleation theory (CNT) and its applicability to aggregation in polyglutamine solutions. Using CNT, we re-analyze previously reported kinetic data for the aggregation of polyglutamine, and interpret the results based on the inferences from knowledge of the phase diagrams in Chapter 3.

In Chapter 5, we shift the focus away from polyglutamine to Amyloid-beta (Aβ), which is one of the proteins implicated in Alzheimer's disease pathogenesis. (*215-217*) *In vitro* studies suggest that Aβ should not aggregate at physiological concentrations based on previously measured saturation concentrations (*30-37*). However, the presence of extracellular aggregates of Aβ is one of the major pathological hallmarks of Alzheimer's disease. (*46, 47, 218-223*) We explore the possibility that Aβ can be trafficked into lysosomes of neuronal cells and concentrated to levels known to support aggregation.

Finally, Chapter 6 provides a summary of the thesis, implications of this work, and suggestions for future research to gain a better understanding of protein aggregation and its relevance in disease pathogenesis.

1.6 References

1. Nozaki, Y., and Tanford, C. (1963) The Solubility of Amino Acids and Related Compounds in Aqueous Urea Solutions, *Journal of Biological Chemistry 238*, 4074-4081.
2. Krull, L. H., Wall, J. S., Zobel, H., and Dimler, R. J. (1965) Synthetic Polypeptides Containing Side-Chain Amide Groups: Water-insoluble Polymers*, *Biochemistry 4*, 626-633.
3. Altschuler, E. L., Hud, N. V., Mazrimas, J. A., and Rupp, B. (1997) Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides, *J Pept Res 50*, 73-75.
4. Cook, W. H., and Rose, R. C. (1934) Solubility of Gluten, *Nature 134*, 380-381.
5. Krull, L. H., and Wall, J. S. (1966) Synthetic Polypeptides Containing Side-Chain Amide Groups. Water-Soluble Polymers*, *Biochemistry 5*, 1521-1527.
6. Tracey, M. V. (1966) Solubility-composition relations of gluten proteins, *Nature 211*, 850-851.
7. Fandrich, M. (2007) On the structural definition of amyloid fibrils and other polypeptide aggregates, *Cell Mol Life Sci 64*, 2066-2078.
8. Vinters, H. V. (1990) Amyloid and the central nervous system: the neurobiology, genetics and immunocytochemistry of a process important in neurodegenerative diseases and stroke, *Monogr Pathol*, 55-85.
9. Rigdon, R. H. (1974) Occurrence and association of amyloid with diseases in birds and mammals including man: A review, *Tex Rep Biol Med 32*, 665-682.
10. Koo, E. H., Lansbury, P. T., Jr., and Kelly, J. W. (1999) Amyloid diseases: abnormal protein aggregation in neurodegeneration, *Proc Natl Acad Sci U S A 96*, 9989-9990.
11. Franklin, E. C., Lavie, G., and Zucker-Franklin, D. (1980) Current concepts of the structure and pathogenesis of the amyloid diseases, *Ric Clin Lab 10*, 3-7.
12. He, X. H., Lin, F., and Qin, Z. H. (2010) Current understanding on the pathogenesis of polyglutamine diseases, *Neurosci Bull 26*, 247-256.
13. Koshy, B. T., and Zoghbi, H. Y. (1997) The CAG/polyglutamine tract diseases: gene products and molecular pathogenesis, *Brain Pathol 7*, 927-942.
14. Rubinsztein, D. C., Wyttenbach, A., and Rankin, J. (1999) Intracellular inclusions, pathological markers in diseases caused by expanded polyglutamine tracts?, *J Med Genet 36*, 265-270.
15. Saudou, F., Devys, D., Trottier, Y., Imbert, G., Stoeckel, M. E., Brice, A., and Mandel, J. L. (1996) Polyglutamine expansions and neurodegenerative diseases, *Cold Spring Harb Symp Quant Biol 61*, 639-647.
16. Perutz, M. F. (1999) Glutamine repeats and neurodegenerative diseases: molecular aspects, *Trends Biochem Sci 24*, 58-63.
17. Perutz, M. F. (1996) Glutamine repeats and inherited neurodegenerative diseases: molecular aspects, *Curr Opin Struct Biol 6*, 848-858.
18. Bates, G. P., Mangiarini, L., Wanker, E. E., and Davies, S. W. (1998) Polyglutamine expansion and Huntington's disease, *Biochem Soc Trans 26*, 471-475.

19. Trottier, Y., Lutz, Y., Stevanin, G., Imbert, G., Devys, D., Cancel, G., Saudou, F., Weber, C., David, G., Tora, L., Agid, Y., Brice, A., and Mandel, J.-L. (1995) Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias, *Nature 378*, 403-406.

20. Gatchel, J. R., and Zoghbi, H. Y. (2005) Diseases of unstable repeat expansion: Mechanisms and common principles, *Nature Reviews Genetics 6*, 743-755.

21. Lieberman, A. P., and Fischbeck, K. H. (2000) Triplet repeat expansion in neuromuscular disease, *Muscle & Nerve 23*, 843-850.

22. MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., MacFarlane, H., Jenkins, B., Anderson, M. A., Wexler, N. S., Gusella, J. F., Bates, G. P., Baxendale, S., Hummerich, H., and Kirby, S. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes, *Cell 72*, 971.

23. Masino, L., and Pastore, A. (2002) Glutarnine repeats: structural hypotheses and neurodegeneration, *Biochemical Society Transactions 30*, 548-551.

24. Paulson, H. L., Bonini, N. M., and Roth, K. A. (2000) Polyglutamine disease and neuronal cell death, *Proceedings of the National Academy of Sciences of the United States of America 97*, 12957-12958.

25. Riley, B. E., and Orr, H. T. (2006) Polyglutamine neurodegenerative diseases and regulation of transcription: assembling the puzzle, *Genes & Development 20*, 2183-2192.

26. Ross, C. A. (1995) When more is less: Pathogenesis of glutamine repeat neurodegenerative diseases, *Neuron 15*, 493-496.

27. Ross, C. A., and Poirier, M. A. (2004) Protein aggregation and neurodegenerative disease., *Nat Med 10 Suppl*, S10-17.

28. Skipper, M. (2008) Human disease: A two-pronged model of polyglutamine disease, *Nature Reviews Genetics 9*.

29. Stott, K., Blackburn, J. M., Butler, P. J., and Perutz, M. (1995) Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases, *Proc Natl Acad Sci U S A 92*, 6509-6513.

30. Sengupta, P., Garai, K., Sahoo, B., Shi, Y., Callaway, D. J., and Maiti, S. (2003) The amyloid beta peptide (Abeta(1-40)) is thermodynamically soluble at physiological concentrations, *Biochemistry 42*, 10506-10513.

31. Lomakin, A., Teplow, D. B., Kirschner, D. A., and Benedek, G. B. (1997) Kinetic theory of fibrillogenesis of amyloid beta-protein, *Proceedings of the National Academy of Sciences of the United States of America 94*, 7942-7947.

32. Harper, J. D., and Lansbury, P. T., Jr. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins, *Annu Rev Biochem 66*, 385-407.

33. Strozyk, D., Blennow, K., White, L. R., and Launer, L. J. (2003) CSF Abeta 42 levels correlate with amyloid-neuropathology in a population-based autopsy study, *Neurology 60*, 652-656.

34. Seubert, P., Vigo-Pelfrey, C., Esch, F., Lee, M., Dovey, H., Davis, D., Sinha, S., Schlossmacher, M., Whaley, J., Swindlehurst, C., and et al. (1992) Isolation and

quantification of soluble Alzheimer's beta-peptide from biological fluids, *Nature 359*, 325-327.

35. Gustafson, D. R., Skoog, I., Rosengren, L., Zetterberg, H., and Blennow, K. (2007) Cerebrospinal fluid beta-amyloid 1-42 concentration may predict cognitive decline in older women, *Journal of neurology, neurosurgery, and psychiatry 78*, 461-464.

36. Galasko, D., Chang, L., Motter, R., Clark, C. M., Kaye, J., Knopman, D., Thomas, R., Kholodenko, D., Schenk, D., Lieberburg, I., Miller, B., Green, R., Basherad, R., Kertiles, L., Boss, M. A., and Seubert, P. (1998) High cerebrospinal fluid tau and low amyloid beta42 levels in the clinical diagnosis of Alzheimer disease and relation to apolipoprotein E genotype, *Archives of neurology 55*, 937-945.

37. Andreasen, N., Hesse, C., Davidsson, P., Minthon, L., Wallin, A., Winblad, B., Vanderstichele, H., Vanmechelen, E., and Blennow, K. (1999) Cerebrospinal fluid beta-amyloid(1-42) in Alzheimer disease: differences between early- and late-onset Alzheimer disease and stability during the course of disease, *Archives of neurology 56*, 673-680.

38. Pepys, M. B. (2006) Amyloidosis, *Annu. Rev. Med. 57*, 223-241.

39. Chiti, F., and Dobson, C. M. (2006) Protein Misfolding, Functional Amyloid, and Human Disease, *Annual Review of Biochemistry 75*, 333-366.

40. Taylor, J. P., Hardy, J., and Fischbeck, K. H. (2002) Toxic proteins in neurodegenerative disease, *Science 296*, 1991-1995.

41. Bucciantini, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature 416*, 507-511.

42. Cohen, F. E. (1999) Protein misfolding and prion diseases, *J. Mol. Biol. 293*, 313-320.

43. Carrell, R. W., and Lomas, D. A. (1997) Conformational disease, *Lancet 350*, 134-138.

44. Ross, C. A., and Poirier, M. A. (2004) Protein aggregation and neurodegenerative disease, *Nature Med. 10*, S10-17.

45. Selkoe, D. J. (2008) Biochemistry and molecular biology of amyloid beta-protein and the mechanism of Alzheimer's disease, *Handb Clin Neurol 89*, 245-260.

46. Rossor, M. N. (1993) Molecular pathology of Alzheimer's disease, *J Neurol Neurosurg Psychiatry 56*, 583-586.

47. Blass, J. P., Ko, L., and Wisniewski, H. M. (1991) Pathology of Alzheimer's disease, *Psychiatr Clin North Am 14*, 397-420.

48. Muller-Hill, B., and Beyreuther, K. (1989) Molecular biology of Alzheimer's disease, *Annu Rev Biochem 58*, 287-307.

49. Takahashi, T., Katada, S., and Onodera, O. (2010) Polyglutamine Diseases: Where does Toxicity Come from? What is Toxicity? Where are We Going?, *Journal of Molecular Cell Biology 2*, 180-191.

50. La Spada, A. R., and Taylor, J. P. (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis, *Nat Rev Genet 11*, 247-258.

51. Swami, M., Hendricks, A. E., Gillis, T., Massood, T., Mysore, J., Myers, R. H., and Wheeler, V. C. (2009) Somatic expansion of the Huntington's disease CAG

repeat in the brain is associated with an earlier age of disease onset, *Hum Mol Genet 18*, 3039-3047.

52. Mirkin, S. M. (2006) DNA structures, repeat expansions and human hereditary disorders, *Curr Opin Struct Biol 16*, 351-358.

53. Bauer, P., Laccone, F., Rolfs, A., Wullner, U., Bosch, S., Peters, H., Liebscher, S., Scheible, M., Epplen, J. T., Weber, B. H., Holinski-Feder, E., Weirich-Schwaiger, H., Morris-Rosendahl, D. J., Andrich, J., and Riess, O. (2004) Trinucleotide repeat expansion in SCA17/TBP in white patients with Huntington's disease-like phenotype, *J Med Genet 41*, 230-232.

54. Margolis, R. L., O'Hearn, E., Rosenblatt, A., Willour, V., Holmes, S. E., Franz, M. L., Callahan, C., Hwang, H. S., Troncoso, J. C., and Ross, C. A. (2001) A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion, *Ann Neurol 50*, 373-380.

55. Holmes, S. E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H. S., Ingersoll-Ashworth, R. G., Fleisher, A., Stevanin, G., Brice, A., Potter, N. T., Ross, C. A., and Margolis, R. L. (2001) A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2, *Nat Genet 29*, 377-378.

56. Cummings, C. J., and Zoghbi, H. Y. (2000) Trinucleotide repeats: mechanisms and pathophysiology, *Annu Rev Genomics Hum Genet 1*, 281-328.

57. Margolis, R. L., McInnis, M. G., Rosenblatt, A., and Ross, C. A. (1999) Trinucleotide repeat expansion and neuropsychiatric disease, *Arch Gen Psychiatry 56*, 1019-1031.

58. Hashida, H., Goto, J., Kurisaki, H., Mizusawa, H., and Kanazawa, I. (1997) Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and spinocerebellar ataxia type 1, *Ann Neurol 41*, 505-511.

59. Trinka, E., Luthringshausen, G., Ladurner, G., Waigell-Weber, M., and Spiegel, R. (1996) Correlations between triplet repeat expansion and clinical features in Huntington's disease, *Arch Neurol 53*, 714-715.

60. Jou, Y.-S., and Myers, R. M. (1995) Evidence from antibody studies that the CAG repeat in the Huntington disease gene is expressed in the protein, *Human Molecular Genetics 4*, 465-469.

61. Cancel, G., Abbas, N., Stevanin, G., Durr, A., Chneiweiss, H., Neri, C., Duyckaerts, C., Penet, C., Cann, H. M., Agid, Y., and et al. (1995) Marked phenotypic heterogeneity associated with expansion of a CAG repeat sequence at the spinocerebellar ataxia 3/Machado-Joseph disease locus, *Am J Hum Genet 57*, 809-816.

62. Brooks, B. P., and Fischbeck, K. H. (1995) Spinal and bulbar muscular atrophy: a trinucleotide-repeat expansion neurodegenerative disease, *Trends Neurosci 18*, 459-461.

63. La Spada, A. R., Paulson, H. L., and Fischbeck, K. H. (1994) Trinucleotide repeat expansion in neurological disease, *Ann Neurol 36*, 814-822.

64. Zoghbi, H. Y., Frontali, M., Orr, H. T., Sandkuijl, L., Cann, H., Sasaki, H., Chamberlain, S., Terrenato, L., and Rich, S. S. (1993) Linkage studies in dominantly inherited ataxias, *Adv Neurol 61*, 133-137.

65. Snell, R. G., MacMillan, J. C., Cheadle, J. P., Fenton, I., Lazarou, L. P., Davies, P., MacDonald, M. E., Gusella, J. F., Harper, P. S., and Shaw, D. J. (1993) Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease, *Nat Genet 4*, 393-397.

66. Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P., and Zoghbi, H. Y. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1, *Nat Genet 4*, 221-226.

67. Myers, R. H., MacDonald, M. E., Koroshetz, W. J., Duyao, M. P., Ambrose, C. M., Taylor, S. A., Barnes, G., Srinidhi, J., Lin, C. S., Whaley, W. L., and et al. (1993) De novo expansion of a (CAG)n repeat in sporadic Huntington's disease, *Nat Genet 5*, 168-173.

68. MacDonald, M. E., Barnes, G., Srinidhi, J., Duyao, M. P., Ambrose, C. M., Myers, R. H., Gray, J., Conneally, P. M., Young, A., Penney, J., and et al. (1993) Gametic but not somatic instability of CAG repeat length in Huntington's disease, *J Med Genet 30*, 982-986.

69. Gusella, J. F., MacDonald, M. E., Ambrose, C. M., and Duyao, M. P. (1993) Molecular genetics of Huntington's disease, *Arch Neurol 50*, 1157-1163.

70. Gusella, J. F., and MacDonald, M. E. (1993) Hunting for Huntington's disease, *Mol Genet Med 3*, 139-158.

71. Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M., Abbott, M., and et al. (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease, *Nat Genet 4*, 387-392.

72. Chung, M. Y., Ranum, L. P., Duvick, L. A., Servadio, A., Zoghbi, H. Y., and Orr, H. T. (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I, *Nat Genet 5*, 254-258.

73. Banfi, S., Chung, M. Y., Kwiatkowski, T. J., Jr., Ranum, L. P., McCall, A. E., Chinault, A. C., Orr, H. T., and Zoghbi, H. Y. (1993) Mapping and cloning of the critical region for the spinocerebellar ataxia type 1 gene (SCA1) in a yeast artificial chromosome contig spanning 1.2 Mb, *Genomics 18*, 627-635.

74. Andrew, S. E., Goldberg, Y. P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., Lin, B., Kalchman, M. A., and et al. (1993) The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease, *Nat Genet 4*, 398-403.

75. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group, *Cell 72*, 971-983.

76. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, K. H. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy, *Nature 352*, 77-79.

77. Petruska, J., Hartenstine, M. J., and Goodman, M. F. (1998) Analysis of Strand Slippage in DNA Polymerase Expansions of CAG/CTG Triplet Repeats Associated with Neurodegenerative Disease, *Journal of Biological Chemistry 273*, 5204-5210.

78. Yamada, M., Sato, T., Tsuji, S., and Takahashi, H. (2008) CAG repeat disorder models and human neuropathology: similarities and differences, *Acta Neuropathologica 115*, 71-86.

79. Schöls, L., Bauer, P., Schmidt, T., Schulte, T., and Riess, O. (2004) Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis, *The Lancet Neurology 3*, 291-304.

80. Yamada, M., Tsuji, S., and Takahashi, H. (2000) Pathology of CAG repeat diseases, *Neuropathology 20*, 319-325.

81. Gusella, J. F., and MacDonald, M. E. (2000) Molecular genetics: Unmasking polyglutamine triggers in neurodegenerative disease, *Nat Rev Neurosci 1*, 109-115.

82. Robitaille, Y., Lopes-Cendes, I., Becher, M., Rouleau, G., and Clark, A. W. (1997) The Neuropathology of CAG Repeat Diseases: Review and Update of Genetic and Molecular Features, *Brain Pathology 7*, 901-926.

83. Reddy, P. S., and Housman, D. E. (1997) The complex pathology of trinucleotide repeats, *Current Opinion in Cell Biology 9*, 364-372.

84. Walker, F. O. (2007) Huntington's disease, *Lancet 369*, 218-228.

85. Burright, E. N., Brent Clark, H., Servadio, A., Matilla, T., Feddersen, R. M., Yunis, W. S., Duvick, L. A., Zoghbi, H. Y., and Orr, H. T. (1995) SCA1 transgenic mice: A model for neurodegeneration caused by an expanded CAG trinucleotide repeat, *Cell 82*, 937-948.

86. Ferrigno, P., and Silver, P. A. (2000) Polyglutamine expansions: proteolysis, chaperones, and the dangers of promiscuity, *Neuron 26*, 9-12.

87. Tarlac, V., and Storey, E. (2003) Role of proteolysis in polyglutamine disorders, *J Neurosci Res 74*, 406-416.

88. Walsh, R., Storey, E., Stefani, D., Kelly, L., and Turnbull, V. (2005) The roles of proteolysis and nuclear localisation in the toxicity of the polyglutamine diseases. A review, *Neurotox Res 7*, 43-57.

89. Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J. L., Fischbeck, K. H., and Pittman, R. N. (1997) Intranuclear Inclusions of Expanded Polyglutamine Protein in Spinocerebellar Ataxia Type 3, *Neuron 19*, 333-344.

90. Hayashi, Y., Kakita, A., Yamada, M., Koide, R., Igarashi, S., Takano, H., Ikeuchi, T., Wakabayashi, K., Egawa, S., Tsuji, S., and Takahashi, H. (1998) Hereditary dentatorubral-pallidoluysian atrophy: detection of widespread ubiquitinated neuronal and glial intranuclear inclusions in the brain, *Acta Neuropathologica 96*, 547-552.

91. Schmidt, T., Lindenberg, K. S., Krebs, A., Schöls, L., Laccone, F., Herms, J., Rechsteiner, M., Riess, O., and Landwehrmeyer, G. B. (2002) Protein surveillance machinery in brains with spinocerebellar ataxia type 3: Redistribution and differential recruitment of 26S proteasome subunits and chaperones to neuronal intranuclear inclusions, *Annals of Neurology 51*, 302-310.

92. Satyal, S. H., Schmidt, E., Kitagawa, K., Sondheimer, N., Lindquist, S., Kramer, J. M., and Morimoto, R. I. (2000) Polyglutamine aggregates alter protein folding homeostasis in Caenorhabditis elegans, *Proceedings of the National Academy of Sciences 97*, 5750-5755.

93.	Yuan, J., and Yankner, B. A. (2000) Apoptosis in the nervous system, *Nature 407*, 802-809.

94.	Steffan, J. S., Kazantsev, A., Spasic-Boskovic, O., Greenwald, M., Zhu, Y.-Z., Gohler, H., Wanker, E. E., Bates, G. P., Housman, D. E., and Thompson, L. M. (2000) The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription, *Proceedings of the National Academy of Sciences 97*, 6763-6768.

95.	Dunah, A. W., Jeong, H., Griffin, A., Kim, Y.-M., Standaert, D. G., Hersch, S. M., Mouradian, M. M., Young, A. B., Tanese, N., and Krainc, D. (2002) Sp1 and TAFII130 Transcriptional Activity Disrupted in Early Huntington's Disease, *Science 296*, 2238-2243.

96.	Schaffar, G., Breuer, P., Boteva, R., Behrends, C., Tzvetkov, N., Strippel, N., Sakahira, H., Siegers, K., Hayer-Hartl, M., and Hartl, F. U. (2004) Cellular Toxicity of Polyglutamine Expansion Proteins: Mechanism of Transcription Factor Deactivation, *Molecular Cell 15*, 95-105.

97.	Sugars, K. L., and Rubinsztein, D. C. (2003) Transcriptional abnormalities in Huntington disease, *Trends in Genetics 19*, 233-238.

98.	Nucifora, F. C., Sasaki, M., Peters, M. F., Huang, H., Cooper, J. K., Yamada, M., Takahashi, H., Tsuji, S., Troncoso, J., Dawson, V. L., Dawson, T. M., and Ross, C. A. (2001) Interference by Huntingtin and Atrophin-1 with CBP-Mediated Transcription Leading to Cellular Toxicity, *Science 291*, 2423-2428.

99.	Yang, W., Dunlap, J. R., Andrews, R. B., and Wetzel, R. (2002) Aggregated polyglutamine peptides delivered to nuclei are toxic to mammalian cells, *Hum. Mol. Genet. 11*, 2905-2917.

100.	Shewry, P. R., and Halford, N. G. (2002) Cereal seed storage proteins: structures, properties and role in grain utilization, *J Exp Bot 53*, 947-958.

101.	Osborne, T. B., and Harris, I. F. (1905) THE CHEMISTRY OF THE PROTEIN-BODIES OF THE WHEAT KERNEL. PART I.—THE PROTEIN SOLUBLE IN ALCOHOL AND ITS GLUTAMINIC ACID CONTENT, *American Journal of Physiology -- Legacy Content 13*, 35-44.

102.	Osborne, T. B., and Nolan, O. L. (1920) DOES GLIADIN CONTAIN AMIDE NITROGEN?, *Journal of Biological Chemistry 43*, 311-316.

103.	(1934) AMIDE NITROGEN IN GERMINATING SEEDS, *Canadian Journal of Research 10*, 430-434.

104.	Stewart, W. D. P. (1967) Nitrogen-Fixing Plants, *Science 158*, 1426-1432.

105.	Herman, E. M., and Larkins, B. A. (1999) Protein Storage Bodies and Vacuoles, *The Plant Cell Online 11*, 601-614.

106.	Kopito, R. R. (2000) Aggresomes, inclusion bodies and protein aggregation, *Trends in Cell Biology 10*, 524-530.

107.	Fink, A. L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid, *Folding and Design 3*, R9-R23.

108.	Shutov, A. D., Bäumlein, H., Blattner, F. R., and Müntz, K. (2003) Storage and mobilization as antagonistic functional constraints on seed storage globulin evolution, *Journal of Experimental Botany 54*, 1645-1654.

109.	Schuppan, D. (2000) Current Concepts of Celiac Disease Pathogenesis, *Gastroenterology 119*, 234-242.

110. McADAM, S. N., and SOLLID, L. M. (2000) Getting to grips with gluten, *Gut 47*, 743-745.

111. Sollid, L. M. (2000) Molecular basis of celiac disease, *Annu. Rev. Immunol. 18*, 53-81.

112. Dicke, W. K., Weijers, H. A., and Van De Kamer, J. H. (1953) Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease, *Acta Paediatr 42*, 34-42.

113. Hausch, F., Shan, L., Santiago, N. A., Gray, G. M., and Khosla, C. (2002) Intestinal digestive resistance of immunodominant gliadin peptides, *American Journal of Physiology - Gastrointestinal and Liver Physiology 283*, G996-G1003.

114. Bethune, M. T., and Khosla, C. (2008) Parallels between Pathogens and Gluten Peptides in Celiac Sprue, *PLoS Pathog 4*, e34.

115. Shan, L., Molberg, Ø., Parrot, I., Hausch, F., Filiz, F., Gray, G. M., Sollid, L. M., and Khosla, C. (2002) Structural Basis for Gluten Intolerance in Celiac Sprue, *Science 297*, 2275-2279.

116. Revedin, A., Aranguren, B., Becattini, R., Longo, L., Marconi, E., Lippi, M. M., Skakun, N., Sinitsyn, A., Spiridonova, E., and Svoboda, J. (2010) Thirty thousand-year-old evidence of plant food processing, *Proceedings of the National Academy of Sciences 107*, 18815-18819.

117. Shewry, P. R., Halford, N. G., Belton, P. S., and Tatham, A. S. (2002) The structure and properties of gluten: an elastic protein from wheat grain, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357*, 133-142.

118. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) Intrinsic disorder in transcription factors, *Biochemistry 45*, 6873-6888.

119. Gerber, H. P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S., and Schaffner, W. (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches, *Science 263*, 808-811.

120. Courey, A. J., Holtzman, D. A., Jackson, S. P., and Tjian, R. (1989) Synergistic activation by the glutamine-rich domains of human transcription factor Sp1, *Cell 59*, 827-836.

121. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat Rev Mol Cell Biol 6*, 197-208.

122. Tjian, R., and Maniatis, T. (1994) Transcriptional activation: a complex puzzle with few easy pieces, *Cell 77*, 5-8.

123. Alberts, B., Wilson, J. H., and Hunt, T. (2008) *Molecular biology of the cell*, 5th ed., Garland Science, New York.

124. Berg, J. M., Tymoczko, J. L., and Stryer, L. (2007) *Biochemistry*, 6th ed., W.H. Freeman, New York.

125. Brivanlou, A. H., and Darnell, J. E., Jr. (2002) Signal transduction and the control of gene expression, *Science 295*, 813-818.

126. Wootton, J. C., and Drummond, M. H. (1989) The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins, *Protein Eng 2*, 535-543.

127. Cha, J. H. (2000) Transcriptional dysregulation in Huntington's disease, *Trends Neurosci 23*, 387-392.

128. Bithell, A., Johnson, R., and Buckley, N. J. (2009) Transcriptional dysregulation of coding and non-coding genes in cellular models of Huntington's disease, *Biochem Soc Trans 37*, 1270-1275.

129. De Deken, R. H., Mortier, A., and Wiame, J. M. (1953) The solubility of wheat gluten, *Biochim Biophys Acta 10*, 488.

130. Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases, *Proc Natl Acad Sci U S A 91*, 5355-5358.

131. Perutz, M. F. (1995) Glutamine repeats as polar zippers: their role in inherited neurodegenerative disease, *Mol Med 1*, 718-721.

132. Perutz, M. F., Finch, J. T., Berriman, J., and Lesk, A. (2002) Amyloid fibers are water-filled nanotubes, *Proc Natl Acad Sci U S A 99*, 5591-5595.

133. Sikorski, P., and Atkins, E. (2004) New Model for Crystalline Polyglutamine Assemblies and Their Connection with Amyloid Fibrils, *Biomacromolecules 6*, 425-432.

134. Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, J. J. W., McFarlane, H. T., Madsen, A. O., Riekel, C., and Eisenberg, D. (2007) Atomic structures of amyloid cross-[bgr] spines reveal varied steric zippers, *Nature 447*, 453-457.

135. Sharma, D., Shinchuk, L. M., Inouye, H., Wetzel, R., and Kirschner, D. A. (2005) Polyglutamine homopolymers having 8–45 residues form slablike β-crystallite assemblies, *Proteins: Structure, Function, and Bioinformatics 61*, 398-411.

136. Chen, S., and Wetzel, R. (2001) Solubilization and disaggregation of polyglutamine peptides, *Protein Sci. 10*, 887-891.

137. Chen, S., Berthelier, V., Hamilton, J. B., O'Nuallain, B., and Wetzel, R. (2002) Amyloid-like features of polyglutamine aggregates and their assembly kinetics, *Biochemistry 41*, 7391-7399.

138. Masino, L., Kelly, G., Leonard, K., Trottier, Y., and Pastore, A. (2002) Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins, *FEBS Letters 513*, 267-272.

139. Altschuler, E. L., Hud, N. V., Mazrimas, J. A., and Rupp, B. (2000) Structure of polyglutamine, *FEBS Letters 472*, 166-167.

140. Wang, X., Vitalis, A., Wyczalkowski, M. A., and Pappu, R. V. (2006) Characterizing the conformational ensemble of monomeric polyglutamine, *Proteins 63*, 297-311.

141. Vitalis, A., Wang, X., and Pappu, R. V. (2007) Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories, *Biophys J 93*, 1923-1937.

142. Vitalis, A., Wang, X., and Pappu, R. V. (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization, *J Mol Biol 384*, 279-297.

143. Chen, S., Ferrone, F. A., and Wetzel, R. (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation, *Proc Natl Acad Sci U S A 99*, 11884-11889.

144. Reissenweber, N. J., and Decaro, J. (1969) Dichroism with congo red. A specific test for amyloid?, *Virchows Arch A Pathol Pathol Anat 347*, 254-259.

145. Romhanyi, G. (1971) Selective differentiation between amyloid and connective tissue structures based on the collagen specific topo-optical staining reaction with congo red, *Virchows Arch A Pathol Pathol Anat 354*, 209-222.

146. Rogers, D. R. (1965) Screening for Amyloid with the Thioflavin-T Fluorescent Method, *Am J Clin Pathol 44*, 59-61.

147. Saeed, S. M., and Fine, G. (1967) Thioflavin-T for amyloid detection, *Am J Clin Pathol 47*, 588-593.

148. Khurana, R., Uversky, V. N., Nielsen, L., and Fink, A. L. (2001) Is Congo red an amyloid-specific dye?, *J Biol Chem 276*, 22715-22721.

149. Morimoto, K., Kawabata, K., Kunii, S., Hamano, K., Saito, T., and Tonomura, B. (2009) Characterization of type I collagen fibril formation using thioflavin T fluorescent dye, *J Biochem 145*, 677-684.

150. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction., *Proc Natl Acad Sci U S A 102*, 15400-15405.

151. Chen, S., Ferrone, F. A., and Wetzel, R. (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation., *Proc Natl Acad Sci U S A 99*, 11884-11889.

152. Chen, S., Berthelier, V., Hamilton, J. B., O'Nuallain, B., and Wetzel, R. (2002) Amyloid-like features of polyglutamine aggregates and their assembly kinetics, *Biochemistry 41*, 7391-7399.

153. Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001) Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity, *J Mol Biol 311*, 173-182.

154. Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G. P., Lehrach, H., and Wanker, E. E. (1999) Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology, *Proc Natl Acad Sci U S A 96*, 4604-4609.

155. Sipe, J. D., and Cohen, A. S. (2000) Review: history of the amyloid fibril, *J Struct Biol 130*, 88-98.

156. Krebs, M. R., Domike, K. R., and Donald, A. M. (2009) Protein aggregation: more than just fibrils, *Biochem Soc Trans 37*, 682-686.

157. Novitskaya, V., Bocharova, O. V., Bronstein, I., and Baskakov, I. V. (2006) Amyloid fibrils of mammalian prion protein are highly toxic to cultured cells and primary neurons, *J Biol Chem 281*, 13828-13836.

158. Lorenzo, A., and Yankner, B. A. (1994) Beta-amyloid neurotoxicity requires fibril formation and is inhibited by congo red, *Proc Natl Acad Sci U S A 91*, 12243-12247.

159. Lorenzo, A., Razzaboni, B., Weir, G. C., and Yankner, B. A. (1994) Pancreatic islet cell toxicity of amylin associated with type-2 diabetes mellitus, *Nature 368*, 756-760.

160. LeVine, H., 3rd. (1993) Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution, *Protein Sci 2*, 404-410.

161. Oosawa, F., and Kasai, M. (1962) A theory of linear and helical aggregations of macromolecules, *J Mol Biol 4*, 10-21.

162. Kasai, M., Asakura, S., and Oosawa, F. (1962) The cooperative nature of G-F transformation of actin, *Biochim Biophys Acta 57*, 22-31.

163. Frieden, C., and Goddette, D. W. (1983) Polymerization of actin and actin-like systems: evaluation of the time course of polymerization in relation to the mechanism, *Biochemistry 22*, 5836-5843.

164. Frieden, C. (1985) Actin and tubulin polymerization: the use of kinetic methods to determine mechanism, *Annu Rev Biophys Biophys Chem 14*, 189-210.

165. Ferrone, F. A. (1993) The polymerization of sickle hemoglobin in solutions and cells, *Experientia 49*, 110-117.

166. Roberts, C. J. (2007) Non-native protein aggregation kinetics, *Biotechnol Bioeng 98*, 927-938.

167. Powers, E. T., and Powers, D. L. (2006) The kinetics of nucleated polymerizations at high concentrations: amyloid fibril formation near and above the "supercritical concentration", *Biophys J 91*, 122-132.

168. Frieden, C. (2007) Protein aggregation processes: In search of the mechanism, *Protein Sci 16*, 2334-2344.

169. Bernacki, J. P., and Murphy, R. M. (2009) Model discrimination and mechanistic interpretation of kinetic data in protein aggregation studies, *Biophys J 96*, 2871-2887.

170. Morris, A. M., Watzky, M. A., and Finke, R. G. (2009) Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature, *Biochim Biophys Acta 1794*, 375-397.

171. Morris, A. M., Watzky, M. A., Agar, J. N., and Finke, R. G. (2008) Fitting neurological protein aggregation kinetic data via a 2-step, minimal/"Ockham's razor" model: the Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth, *Biochemistry 47*, 2413-2427.

172. Ferrone, F. (1999) Analysis of protein aggregation kinetics, *Methods Enzymol 309*, 256-274.

173. Oosawa, F., and Asakura, S. (1975) *Thermodynamics of the polymerization of protein* Academic Press.

174. De Greef, T. F., Smulders, M. M., Wolffs, M., Schenning, A. P., Sijbesma, R. P., and Meijer, E. W. (2009) Supramolecular polymerization, *Chem Rev 109*, 5687-5754.

175. Ciferri, A. (2005) *Supramolecular Polymers*, CRC Press.

176. O'Nuallain, B., Williams, A. D., Westermark, P., and Wetzel, R. (2004) Seeding specificity in amyloid growth induced by heterologous fibrils, *J Biol Chem 279*, 17490-17499.

177. Krebs, M. R., Wilkins, D. K., Chung, E. W., Pitkeathly, M. C., Chamberlain, A. K., Zurdo, J., Robinson, C. V., and Dobson, C. M. (2000) Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the beta-domain, *J Mol Biol 300*, 541-549.

178. Jarrett, J. T., and Lansbury, P. T., Jr. (1993) Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie?, *Cell 73*, 1055-1058.

179. Come, J. H., Fraser, P. E., and Lansbury, P. T., Jr. (1993) A kinetic model for amyloid formation in the prion diseases: importance of seeding, *Proc Natl Acad Sci U S A 90*, 5959-5963.

180. Xue, W. F., Homans, S. W., and Radford, S. E. (2008) Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly, *Proc Natl Acad Sci U S A 105*, 8926-8931.

181. Li, Y., and Roberts, C. J. (2009) Lumry-Eyring nucleated-polymerization model of protein aggregation kinetics. 2. Competing growth via condensation and chain polymerization, *J Phys Chem B 113*, 7020-7032.

182. Andrews, J. M., and Roberts, C. J. (2007) A Lumry-Eyring nucleated polymerization model of protein aggregation kinetics: 1. Aggregation with pre-equilibrated unfolding, *J Phys Chem B 111*, 7897-7913.

183. Harper, J. D., Wong, S. S., Lieber, C. M., and Lansbury, P. T., Jr. (1999) Assembly of A beta amyloid protofibrils: an in vitro model for a possible early event in Alzheimer's disease, *Biochemistry 38*, 8972-8980.

184. Rochet, J. C., and Lansbury, P. T., Jr. (2000) Amyloid fibrillogenesis: themes and variations, *Curr Opin Struct Biol 10*, 60-68.

185. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction, *Proc Natl Acad Sci U S A 102*, 15400-15405.

186. Vitalis, A., and Pappu, R. V. (2011) Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides, *Biophys Chem*.

187. Vitalis, A., Lyle, N., and Pappu, R. V. (2009) Thermodynamics of beta-sheet formation in polyglutamine, *Biophys J 97*, 303-311.

188. Lee, C. C., Walters, R. H., and Murphy, R. M. (2007) Reconsidering the mechanism of polyglutamine peptide aggregation., *Biochemistry 46*, 12810-12820.

189. Perczel, A., Hudaky, P., and Palfi, V. K. (2007) Dead-end street of protein folding: thermodynamic rationale of amyloid fibril formation, *J Am Chem Soc 129*, 14959-14965.

190. Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease, *Annu Rev Biochem 75*, 333-366.

191. Chiti, F., and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions, *Nat Chem Biol 5*, 15-22.

192. Flory, P. J. (1953) *Principles of Polymer Chemistry*, Cornell University Press, Ithaca and London.

193. Ganazzoli, F., Raos, G., and Allegra, G. (1999) Polymer association in poor solvents: from monomolecular micelles to clusters of chains and phase separation, *Macromolecular Theory And Simulations 8*, 65-84.

194. Grosberg, A. Y., and Khokhlov, A. R. (1997) *Statistical Physics of Macromolecules (AIP Series in Polymers and Complex Materials)*, 1 ed., AIP Press, New York.

195. Grosberg, A. Y., and Kuznetsov, D. V. (1992) Phase-Separation Of Polymer-Solutions And Interactions Of Globules, *Journal De Physique Ii 2*, 1327-1339.

196.   Grosberg, A. Y., and Kuznetsov, D. V. (1992) Quantitative Theory Of The Globule-To-Coil Transition .4. Comparison Of Theoretical Results With Experimental-Data, *Macromolecules 25*, 1996-2003.

197.   Grosberg, A. Y., and Kuznetsov, D. V. (1992) Quantitative Theory Of The Globule-To-Coil Transition .3. Globule Globule Interaction And Polymer-Solution Binodal And Spinodal Curves In The Globular Range, *Macromolecules 25*, 1991-1995.

198.   Grosberg, A. Y., and Kuznetsov, D. V. (1992) Quantitative Theory Of The Globule-To-Coil Transition .2. Density Density Correlation In A Globule And The Hydrodynamic Radius Of A Macromolecule, *Macromolecules 25*, 1980-1990.

199.   Grosberg, A. Y., and Kuznetsov, D. V. (1992) Quantitative Theory Of The Globule-To-Coil Transition .1. Link Density Distribution In A Globule And Its Radius Of Gyration, *Macromolecules 25*, 1970-1979.

200.   Pappu, R. V., Wang, X., Vitalis, A., and Crick, S. L. (2008) A polymer physics perspective on driving forces and mechanisms for protein aggregation, *Arch Biochem Biophys 469*, 132-141.

201.   Rubinstein, M., and Colby, R. H. (2003) *Polymer Physics*, Oxford University Press, Oxford and New York.

202.   Raos, G., and Allegra, G. (1996) Chain interactions in poor-solvent polymer solutions: Equilibrium and nonequilibrium aspects, *Macromolecules 29*, 6663-6670.

203.   Raos, G., and Allegra, G. (1996) Chain collapse and phase separation in poor-solvent polymer solutions: A unified molecular description, *Journal Of Chemical Physics 104*, 1626-1645.

204.   Raos, G., and Allegra, G. (1997) Macromolecular clusters in poor-solvent polymer solutions, *Journal Of Chemical Physics 107*, 6479-6490.

205.   Bondos, S. E. (2006) Methods for Measuring Protein Aggregation, *Current Analytical Chemistry 2*, 157-170.

206.   Vernaglia, B. A., Huang, J., and Clark, E. D. (2004) Guanidine hydrochloride can induce amyloid fibril formation from hen egg-white lysozyme, *Biomacromolecules 5*, 1362-1370.

207.   Panayiotou, C., and Sanchez, I. C. (1991) Statistical Thermodynamics of Associated Polymer-Solutions, *Macromolecules 24*, 6231-6237.

208.   Kammer, H. W., Inoue, T., and Ougizawa, T. (1989) Upper and Lower Critical Solution Temperature Behavior in Polymer Blends and Its Thermodynamic Interpretation, *Polymer 30*, 888-892.

209.   Ougizawa, T., Inoue, T., and Kammer, H. W. (1985) Upper and Lower Critical Solution Temperature Behavior in Polymer Blends, *Macromolecules 18*, 2089-2092.

210.   Tenbrinke, G., and Karasz, F. E. (1984) Lower Critical Solution Temperature Behavior in Polymer Blends - Compressibility and Directional-Specific Interactions, *Macromolecules 17*, 815-820.

211.   Qian, C. B., Mumby, S. J., and Eichinger, B. E. (1991) Phase-Diagrams of Binary Polymer-Solutions and Blends, *Macromolecules 24*, 1655-1661.

212. Lu, J., Carpenter, K., Li, R.-J., Wang, X.-J., and Ching, C.-B. (2004) Cloud-point temperature and liquid-liquid phase separation of supersaturated lysozyme solution, *Biophysical Chemistry 109*, 105-112.

213. Mumby, S. J., and Sher, P. (1994) Determination of Chi from Liquid-Liquid Phase Data and the Computation of Phase-Diagrams for Quasi-Binary Polymer-Solutions and Blends, *Macromolecules 27*, 689-694.

214. Chuang, J., A. Y., G., and T., T. (2000) Topological repulsion between polymer globules, *The Journal of Chemical Physics 112*, 6434-6442.

215. Knauer, M. F., Soreghan, B., Burdick, D., Kosmoski, J., and Glabe, C. G. (1992) Intracellular accumulation and resistance to degradation of the Alzheimer amyloid A4/beta protein, *Proc Natl Acad Sci U S A 89*, 7437-7441.

216. Halverson, K., Fraser, P. E., Kirschner, D. A., and Lansbury, P. T., Jr. (1990) Molecular determinants of amyloid deposition in Alzheimer's disease: conformational studies of synthetic beta-protein fragments, *Biochemistry 29*, 2639-2644.

217. Glenner, G. G., and Wong, C. W. (1984) Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein, *Biochem Biophys Res Commun 120*, 885-890.

218. Behrouz, N., Defossez, A., Delacourte, A., and Mazzuca, M. (1991) The immunohistochemical evidence of amyloid diffuse deposits as a pathological hallmark in Alzheimer's disease, *J Gerontol 46*, B209-212.

219. Anderton, B. H. (1987) Alzheimer's disease. Progress in molecular pathology, *Nature 325*, 658-659.

220. Hyman, B. T., Van Hoesen, G. W., Damasio, A. R., and Barnes, C. L. (1984) Alzheimer's disease: cell-specific pathology isolates the hippocampal formation, *Science 225*, 1168-1170.

221. Mann, D. M., Neary, D., Yates, P. O., Lincoln, J., Snowden, J. S., and Stanworth, P. (1981) Neurofibrillary pathology and protein synthetic capability in nerve cells in Alzheimer's disease, *Neuropathol Appl Neurobiol 7*, 37-47.

222. McKee, A. C., Kosik, K. S., and Kowall, N. W. (1991) Neuritic pathology and dementia in Alzheimer's disease, *Ann Neurol 30*, 156-165.

223. Van Hoesen, G. W., Hyman, B. T., and Damasio, A. R. (1986) Cell-specific pathology in neural systems of the temporal lobe in Alzheimer's disease, *Prog Brain Res 70*, 321-335.

Chapter 2: Polyglutamine forms compact globules in aqueous solutions

2.1 Preamble

This chapter is based on work which was completed and published in 2006. (*1*) This story has continued to evolve since that time, and our lab and others have contributed. This chapter will be introduced and presented in 2.2-2.4 as it was in 2006. In addition to our original discussion, we will also discuss the results in 2.6 in light of new information gathered since the original manuscript was published.

This work was originally motivated by simulations in the Pappu lab which suggested that polyglutamine molecules form collapsed structures in solution with seemingly no preference for the formation of a dominant secondary structure. (*2*) This was consistent with previously published data suggesting that polyglutamine molecules are devoid of secondary structure. (*3-7*) However, there was no experimental evidence available regarding the global shapes and sizes of these molecules. The goal of this work was to provide this information and contribute to a better understanding of the solution behavior of polyglutamine molecules.

2.2 Introduction

Polyglutamine stretches are known to aggregate in aqueous solution. (*4, 8, 9*) This is surprising considering that polyglutamine is a concatenation of the polar amino acid glutamine. Circular dichroism (CD) and nuclear magnetic resonance (NMR) data indicate that monomeric polyglutamine sequences prefer the random coil state under physiological conditions. (*3-7*) As chain length increases, there is no obvious change in the ensemble averaged solution "structure" of polyglutamine peptides (*5-7*). The observation that monomeric polyglutamine molecules are essentially devoid of structure led to interpretations that polyglutamine is similar to a fully denatured protein and that perhaps the transition to a pathogenic form is similar to an unfavorable folding event. (*6, 10*)

A problem with CD and NMR data is that these modalities probe local structural information. Neither of these probes provides any information about global shapes and sizes adopted by individual polyglutamine molecules. Our focus is on a complete description of the aqueous solution "structure" of monomeric polyglutamine.

In studies based on molecular dynamics simulations of monomeric polyglutamine it was shown that as chain lengths increase there ought to be an increased probability for sampling compact, roughly spherical geometries. (*2*) These conformations were also devoid of any dominant secondary structures, and are consistent with CD and NMR data. Another measure is necessary to adjudicate between the hypotheses that polyglutamine molecules prefer extended geometries, like fully denatured proteins, or compact

geometries devoid of secondary structure. To distinguish between these two possibilities, we carried out systematic measurements of hydrodynamic properties for monomeric polyglutamine as a function of chain length.

Monomeric polyglutamine is analogous to linear, flexible polymers that have access to conformationally heterogeneous ensembles. In such systems, quantities such as chain size, measured by average radius of gyration ($\langle R_g \rangle$), average hydrodynamic radius ($\langle R_h \rangle$), or average translation diffusion time ($\langle \tau_D \rangle$[1], which is directly proportional to $\langle R_h \rangle$), scale with chain length (N) according to power laws of the form N^ν. (11) $\langle R_g \rangle$ and $\langle R_h \rangle$ are equivalent but not identical measures of chain size. If the polymers are sufficiently long and flexible, ν assumes one of three values: $\nu = 0.59$ for a chain in a good solvent; $\nu = 0.5$ in a theta or indifferent solvent; and $\nu = 0.33$ in a poor solvent.

Solvent quality, as measured by the value of ν, provides quantitative assessment of the balance between chain-solvent and chain-chain interactions for a polymer in a specific environment. It also provides information regarding the preferred sizes and shapes of molecules in solution. In a good solvent ($\nu = 0.59$), there is a marked preference for conformations that promote favorable interactions with the surrounding solvent. (12) Therefore, the ensemble is characterized by large fluctuations, and chains form loosely packed structures with an average preference for prolate ellipsoidal shapes.

---

[1] In polymer physics literature, $\langle \tau_D \rangle$ typically measures the average time it takes for a chain to diffuse across a distance $\sim \langle R_h \rangle$. This is not the case here because $\langle \tau_D \rangle$ quantifies the average time it takes for a fluorescent molecule to diffuse across a confocal beam volume in fluorescence correlation spectroscopy experiments.

(*13*) This appears to be the case for polypeptides in high concentrations of denaturants such as 8 M urea or 6 M GdnCl. (*14*) In a theta solvent ($v = 0.5$), chain-chain and chain-solvent interactions counterbalance exactly and statistical properties of polymers follow those of classical ideal chain models formulated by Flory. (*11, 15, 16*) Finally, in a poor solvent ($v = 0.33$), chain-chain interactions are favored over chain-solvent interactions, and collapsed, roughly spherical conformations dominate the ensemble.

As mentioned previously, UV-CD spectra for monomeric polyglutamine in phosphate buffered saline (PBS) at 37°C resemble those of proteins in strong denaturants. (*3-5*) Does this mean that the ensemble for monomeric polyglutamine in aqueous solutions is akin to those of denatured proteins? This question can be answered by quantifying the scaling of chain size as a function of chain length, which we have done using fluorescence correlation spectroscopy (FCS).

FCS is ideally suited for quantitative studies of systems that are prone to aggregation because we need only nanomolar concentrations of peptide samples to carry out the experiments. (*17-20*) More classical techniques for determining the chain dimensions such as static or dynamic light scattering are not sensitive to the small chain lengths we must use and, therefore, require much higher concentrations of peptide. This leads to aggregation of the peptide, which can make separating contributions to the signal from the monomer alone nearly impossible. At nanomolar concentrations used in FCS the signal from the monomer is not compromised by the possible presence of small amounts of aggregates, which would be easily recognized by large bursts in fluorescence.

Using data collected from an FCS experiment, one can determine the average translational diffusion time, $\langle\tau_D\rangle$, which is directly proportional to the average hydrodynamic radius, $\langle R_h\rangle$. We can then use the scaling behavior of the chain size to determine solvent quality and global structural preferences.

2.3 Methods

2.3.1 Preparation of peptide samples

All peptides were purchased in crude from Yale University's Keck Biotechnology center. The peptides were synthesized using solid-phase synthesis according to the general design Gly-(Gln)$_{<N>}$-Cys*-Lys$_2$. The C-terminal lysine residues were included to increase peptide solubility and a cysteine residue was incorporated in order to label the peptides with fluorophores. The crude forms of peptides were disaggregated (*4, 21*) and purified using reverse-phase HPLC on a preparative C$_3$ reverse-phase column (Agilent) to yield peptides of average length $\langle N\rangle$ = 15, 20, 24, 27, 33, 36, 40, 47, and 53. Identities of the peptides were confirmed using electrospray mass spectrometric analysis. Peptide synthesis yields a narrow distribution of chain lengths around the desired value and purification further narrows this distribution. Nevertheless, purified peptides are mixtures of repeat lengths and $\langle N\rangle$ is therefore a weighted average of these mixtures. The asymmetric horizontal error bars in Figure 2.1 are not true standard deviations. Instead they are meant to denote the range of chain lengths present in each peptide sample.

Upon purification, each peptide sequence was chemically modified via a through-cysteine covalent attachment of the fluorescent dye, AlexaFluor-488-C_5-maleimide (Molecular Probes). Freshly disaggregated peptides were reacted overnight with four-fold excess dye at room temperature in 20 mM Hepes buffer (N-(2-hydroxy ethyl) piperazine –N-(2-ethanesulfonic acid)), pH 8.0 with 10 mM tris-(2-carboxyethyl) phosphine (TCEP) and 5 mM EDTA. Following the labeling reaction, the reaction mixture was lyophilized and subjected to disaggregation. (*4, 21*) All unreacted dye was removed using a size exclusion column with a cutoff of 1400 Daltons (Pierce, Rockford, IL). The unreacted peptide molecules were removed and the desired product was further purified using reverse-phase HPLC (C_3 chromatography column, Agilent). The identities of labeled peptides were confirmed using electrospray mass spectrometry. Purified, labeled, and disaggregated peptides were dissolved in a pH 3.0 trifluoroacetic acid (TFA)–water mixture. Aliquots of 50 μL vials of 1 μM concentrations were made and flash frozen in liquid nitrogen and stored at −80ºC.

Prior to carrying out the FCS measurements, peptide samples were thawed at room temperature. Each sample was diluted to a concentration of 50 nM in Dulbecco's Phosphate Buffered Saline (PBS) at pH 7.4 (8.0g NaCl, 0.2g KCl, 1.15g Na_2PO_4, 0.2g KH_2PO_4, dissolved in pure H_2O, 25°C). 400 μL of this solution was placed in a single well of an 8-Chamber Nunc Lab-Tek 1.0 Borosilicate Coverglass System. One of the wells always contained a 20 nM solution of free Alexa488-maleimide for reference purposes.

2.3.2 FCS Experiments

All measurements were performed on a Confocor II LSM system (Carl Zeiss-Evotec, Jena, Germany) using a 40X water-immersion objective. Data for fluorescence intensity autocorrelation functions were analyzed using the Zeiss Confocor II FCS software. The samples were excited at 488 nm using an Argon laser and emissions were collected in the 505–550 nm range. In all experiments, the laser power was allowed to stabilize for at least thirty minutes before beginning data collection. This was done to minimize any non-linearity during startups, which were monitored by keeping track of photon counts from the free dye. Once the photon counts from the free dye stabilized, the counts from each sample were monitored to account for non-specific adsorption of the peptide to the chamber walls. Typically, photon counts reached a steady-state at peptide concentrations of 10 nM and this was the peptide concentration used in all FCS measurements.

For a given peptide sample in a well, an independent measurement refers to a single 25×25 scan, which corresponds to the collection of FCS data 25 times where the duration for each data collection run was 25 seconds. Each scan yielded a distinct estimate for the diffusion time wherein the autocorrelation curves from all 25 experiments were averaged and the resultant curve was fit using the model shown in equation [1]. We carried out eight different 25×25 scans and obtained eight independent estimates of $\langle \tau_D \rangle$ for each of the nine peptide samples.

$$[1] \quad G(t) = \frac{1}{n} \left(\frac{1 - f_T + f_T \exp\left(\frac{-t}{\tau_T}\right)}{1 - f_T} \right) \left(\frac{1}{(1 + \tau_D)\sqrt{1 + \frac{t}{S^2 \tau_D}}} \right) + 1$$

In equation [1], n is the average number of fluorescent molecules in the beam volume, f_T is the fraction of the triplet state formed per dye molecule, τ_T is the decay constant of the triplet, S is a structure factor which describes the shape of the beam volume, and τ_D is the translational diffusion time. S is a fixed parameter for an independent experiment *i.e.*, for a 25×25 scan. All other parameters in equation [1] were estimated using a Levenberg-Marquardt non-linear least squares fit of the model to observed data. The parameters f_T and τ_T are determined primarily by the photo-physics of the fluorescent dye. As a result, the fitting procedure is deemed to be robust if f_T and τ_T are essentially invariant with chain length.

We are confident that for all chain lengths, the diffusing species is monomeric polyglutamine, rather than a distribution of monomers and small oligomers. This assertion is based on four criteria. First, the concentrations used are orders of magnitude below the estimated concentrations required for aggregation. (*5*) Second, MALDI-TOF mass spectrometry analysis of labeled peptide samples at concentrations higher than those used in the FCS experiments do not show evidence for species other than the monomeric form. Third, if we assume the presence of a second diffusing component, the diffusion times we obtain for this component are considerably smaller than that of the free dye. Fourth, the brightness per molecule in all our measurements is similar to that of the free dye. If labeled molecules formed dimers, then they would appear twice as bright

if dimers were the dominant species or brightness fluctuations would be considerably larger than what we observe and there would be statistically significant outliers from the line of best fit shown in Figure 2.2, but this is not the case.

2.3.3 Data analysis

Data collection yielded eight independent estimates of $\langle \tau_D \rangle$ for each of the nine peptide samples. The goals for data analysis were three-fold: First, to compute the correlation coefficient between $\ln(\langle \tau_D \rangle)$ and $\ln(\langle N \rangle)$; Second, to estimate values for the parameters $\ln(\tau_o)$ and ν using the method of least squares; And third, to assess the goodness of the line of best fit obtained using linear regression analysis.

We used a global analysis based on Monte Carlo bootstrap methods to analyze the data for scaling of $\langle \tau_D \rangle$ as a function of $\langle N \rangle$. (*22*) Each Monte Carlo trial proceeds as follows: A measured value of $\langle \tau_D \rangle$ was drawn at random for each of the nine chain lengths. This leads to the generation of a random dataset. Linear regression analysis was carried out on the dataset to estimate $\ln(\tau_o)$, ν, the residuals from the line of best fit, estimates for standard deviations in prediction errors, and the correlation coefficient between $\ln(\langle \tau_D \rangle)$ and $\ln(\langle N \rangle)$. The procedure of randomly drawing measured data points to create a dataset for linear regression analysis was repeated 5×10^5 times. The results from multiple, independent linear regression analyses were used to compute averages and standard deviations for $\langle \tau_D \rangle$ as shown in Figures 2.1 and 2.2 and to assess error bounds on our estimates for $\ln(\tau_o)$ and ν.

The hydrodynamic radius $\langle R_h \rangle$ is directly proportional to the translational diffusion time $\langle \tau_D \rangle$ and we calculated $\langle R_h \rangle$ from the measured values for $\langle \tau_D \rangle$ using the following prescription. The radial (ω_1) and axial (ω_2) dimensions of the laser beam were identical for each FCS experiment. The former was quantified by measuring the diffusion of the free Alexa dye using the formula $\omega_1 = \sqrt{4D\langle \tau_D \rangle}$. The diffusion constant for the Alexa dye[2] ($D=4.35\times10^{-10}$ms^{-2}) is known (24) and this allowed us to compute the value for ω_1. For each peptide sample, we used the known value for ω_1 (which is fixed) and the measured value of $\langle \tau_D \rangle$ and first calculated the diffusion coefficient D (as shown above) and then used this value to determine $\langle R_h \rangle$ using the Stokes-Einstein relationship given by $D = \frac{kT}{6\pi\eta\langle R_h \rangle}$. Here, $k = 1.38\times10^{-23}$ JK^{-1}, $T = 294.5$K, and η is the viscosity of water which is 9.67×10^{-3} poise at 21.5° C. (25) In using this prescription for computing $\langle R_h \rangle$, we implicitly assume the "non-draining" limit *i.e.*, we are stipulating that there is minimal solvent penetration. This assumption must be regarded with extreme caution and therefore, while the estimated values of $\langle R_h \rangle$ are shown in Figure 2.1, we use only the directly measured quantity namely, $\langle \tau_D \rangle$, to assess solvent quality.

---

[2] The value used in our original 2006 paper was 2.24×10^{-10} ms^{-2} from Pristinski et al. This has no effect on the conclusions, but changes the calculated $\langle \mathbf{R_h} \rangle$ by a factor of ~2.

23. Pristinski, D., Kozlovskaya, V., and Sukhishvili, S. A. (2005) Fluorescence correlation spectroscopy studies of diffusion of a weak polyelectrolyte in aqueous solutions, *J Chem Phys 122*, 14907.

2.3.4 Computer Simulations and Prediction of Hydrodynamic Radii

The simulation results used here were provided by Tim Williamson and Nicholas Lyle from the Pappu lab. All simulations were performed using the ABSINTH implicit solvation model (*26*) and the CAMPARI software engine (*27*) developed in the Pappu lab. The constructs simulated were of the form Ac-Q$_n$-Nme for *n* = 15, 20, 25, 30, 35, and 45, where Ac and Nme are acyl and n-methylamide end-capping groups for the N and C termini, respectively. The simulations were performed as described in Williamson et al. (*28*) We used averages of results from three independent temperature replica-exchange simulations, and the temperature is 25° C. The radius of gyration ($\langle R_g \rangle$) was calculated in ABSINTH using the formula: $\langle R_g^2 \rangle = \langle \frac{1}{N} \sum_{i=1}^{N} (\vec{R}_i - \vec{R}_{cm})^2 \rangle$, where *N* is the number of atoms in the molecule, R_{cm} is the position of the center of mass, and R_i is the position of atom *i*.

The hydrodynamic radius ($\langle R_h \rangle$) was calculated using the program HYDROPRO (*29, 30*) as we did in the work of Mao et al (*31*). Briefly, the structure of Rhodamine 6G was analyzed in HYDROPRO, and the HYDROPRO parameters were calibrated to reproduce the known diffusion coefficient of Rhodamine 6G. Using the calibrated parameters, every 100th frame of each simulation for the polyglutamine constructs described in the previous paragraph was analyzed using HYDROPRO, and $\langle R_h \rangle$ was calculated. The average $\langle R_h \rangle$ was calculated over all analyzed frames from a simulation, and the reported $\langle R_h \rangle$ represents the average from three independent simulations.

2.4 Results

2.4.1 Variation of $\langle \tau_D \rangle$ with chain length (*N)*

Figure 2.1 shows FCS data for the average translational diffusion time $\langle \tau_D \rangle$ (measured in microseconds) as a function of the average chain length $\langle N \rangle$ for the peptide series Gly-(Gln)$_{<N>}$-Cys*-Lys$_2$. A fluorophore is attached to the cysteine residue as described in the methods section. It can be seen that $\langle \tau_D \rangle$ increases monotonically with chain length, which means that there is no abrupt change in the ensemble averaged structure of polyglutamine as the chain length crosses some pathological length threshold ($N > 35$). This observation is consistent with the conclusions of Chen et al. who showed that CD spectra of monomeric polyglutamine peptides are insensitive to variations in chain length. (*5)*
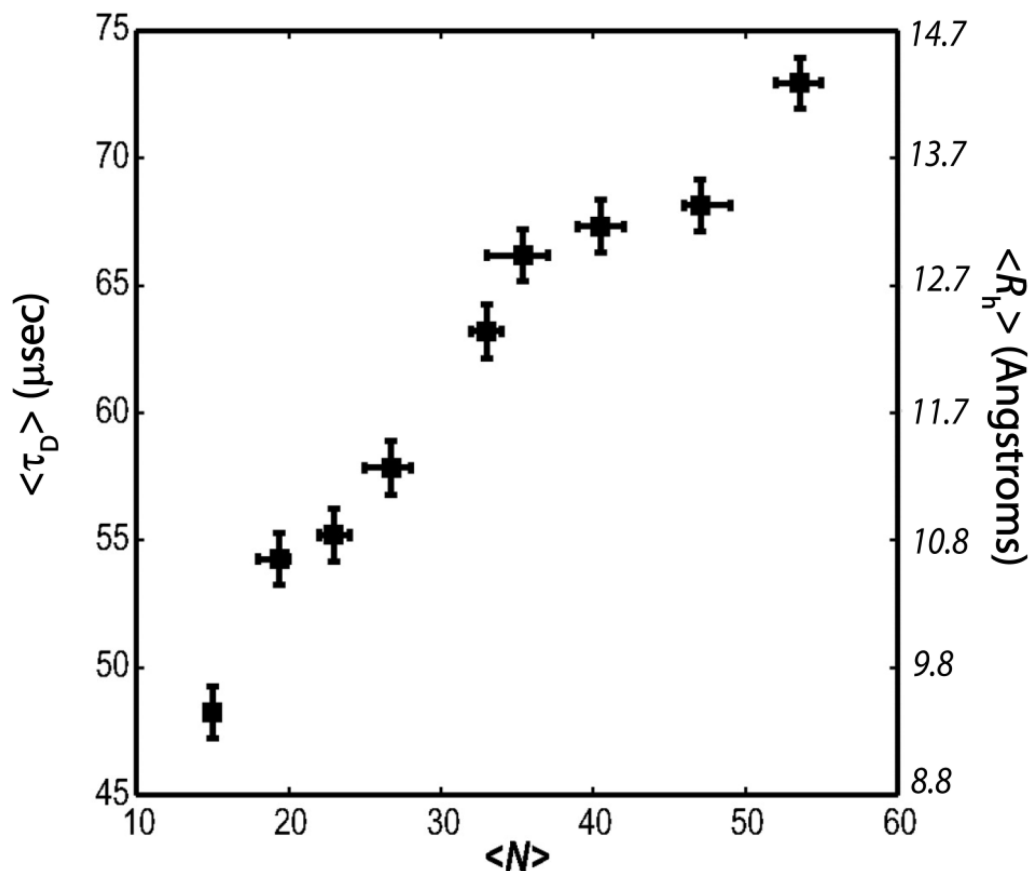
Figure 2.1 Variation of $\langle \tau_D \rangle$ with $\langle N \rangle$, the average number of glutamine residues in the peptide series Gly-(Gln)$_{\langle N \rangle}$-Cys*-(Lys)$_2$. Horizontal error bars are asymmetric because they are not true error bars. Instead, they are meant to convey the range of chain lengths within each peptide sample. The ordinate labels shown on in italics on the right are estimates for the hydrodynamic radius ($\langle R_h \rangle$) in angstroms, which were calculated from measured values for $\langle \tau_D \rangle$ using the prescription described in the methods section.

2.4.2 Linear regression analysis

If $\langle \tau_D \rangle = \tau_o \langle N \rangle^\nu$, then it is also true that $\ln\langle \tau_D \rangle = \ln\tau_o + \nu\ln\langle N \rangle$. The computed linear correlation coefficient between $\ln(\langle \tau_D \rangle)$ and $\ln(\langle N \rangle)$ is 0.961 with a p-value of 8×10^{-5}. This indicates that our data support the hypothesis for the existence of statistically significant linear correlation between $\ln(\langle \tau_D \rangle)$ and $\ln(\langle N \rangle)$.

Figure 2.2 shows results from linear regression analysis. Here, we plot FCS data for $\ln(\langle \tau_D \rangle)$ versus $\ln(\langle N \rangle)$, the line of best fit to the data based on values for ν and $\ln(\tau_o)$ obtained from linear least squares fits to FCS data, and the 95% confidence intervals. The parameters for the line of best fit to the data for $\ln(\langle \tau_D \rangle)$ versus $\ln(\langle N \rangle)$ are a slope ν and intercept $\ln(\tau_o)$ of 0.32 ± 0.02 and 3.04 ± 0.08, respectively. Goodness of fit is assessed by the observation that all points lie within the 95% confidence interval. The norm of the residuals from the line of best fit is 0.02.
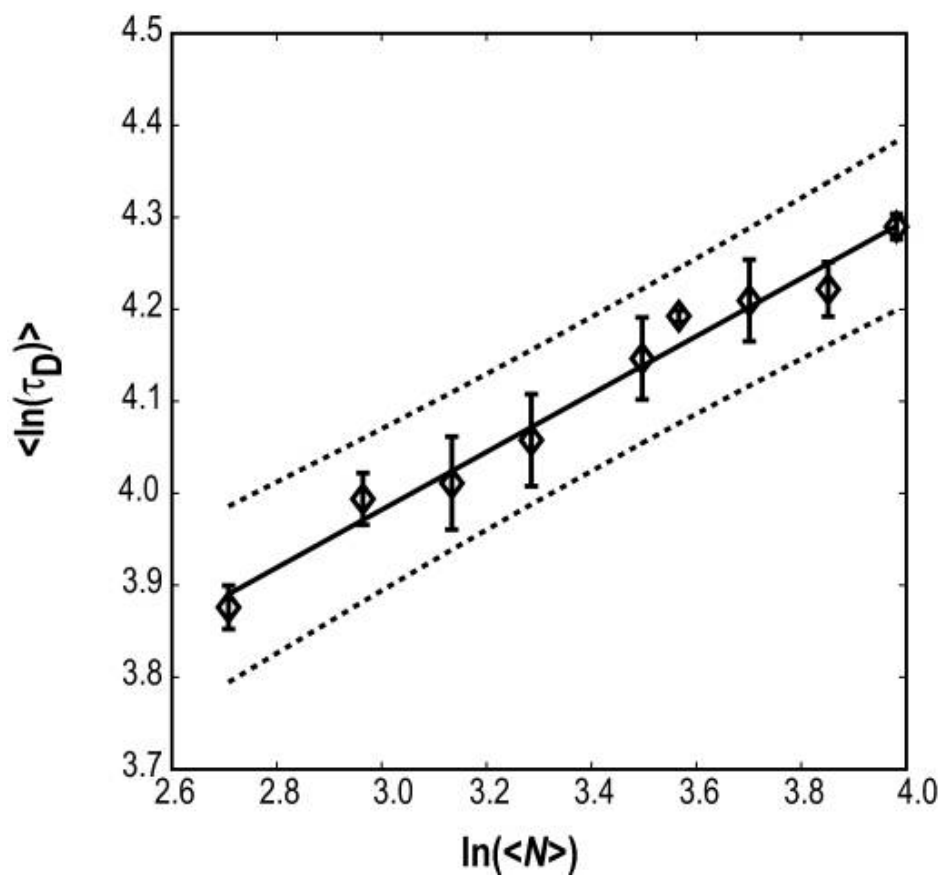
Figure 2.2 Plot of $\ln(\langle \tau_D \rangle)$ against $\ln(\langle N \rangle)$**.** The solid line is the line of best fit to FCS data shown as open diamonds and the dotted lines represent the 95% confidence intervals. Error bars are standard errors in our estimate for the mean value of $\ln(\langle \tau_D \rangle)$, which is obtained from Monte Carlo bootstrap analysis (see methods section).

Based on the value obtained for ν, we conclude that in aqueous solvents, individual polyglutamine molecules behave like chains in a poor solvent. The implication is that these systems prefer collapsed, roughly spherical geometries in aqueous environments.

2.4.3 Comparison to Simulation Results

We have had success calculating hydrodynamic properties from atomistic simulations using a package known as HYDROPRO (*29, 30*) to post-analyze simulation data. (*31*) Mao et al. obtained quantitative agreement between diffusion times measured using FCS and diffusion times calculated using HYDROPRO. Here, we have used HYDROPRO to calculate the hydrodynamic radius of homopolymeric polyglutamine of chain lengths 15, 20, 25, 30, 35, and 45, which were previously simulated in the Pappu lab. This is plotted against experimentally determined values of $\langle R_h \rangle$ in Figure 2.3.
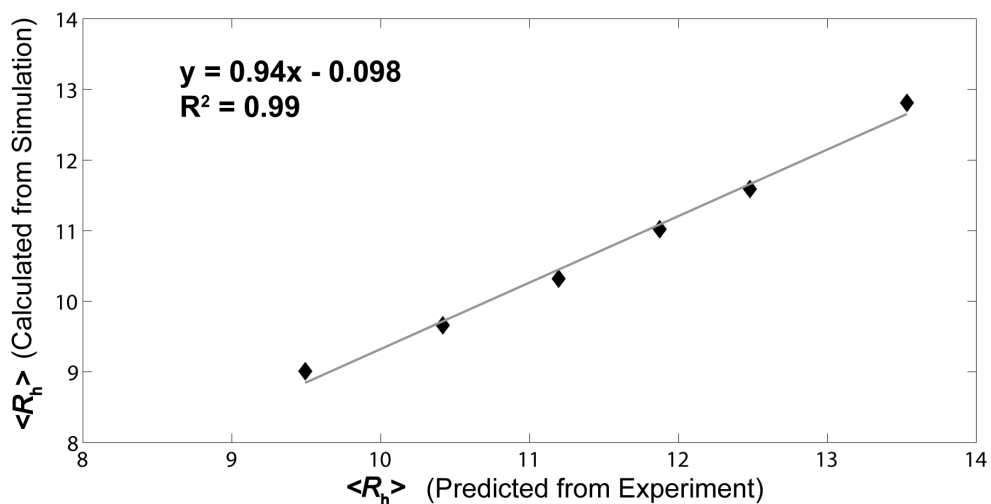
Figure 2.3 Plot of the predicted hydrodynamic radius ($\langle R_h \rangle$) from experiment versus the $\langle R_h \rangle$ of Q_N (N = 15, 20, 25, 30, 35, and 45) calculated from computer simulation data. Because the experiments were not performed on the same chain lengths as the simulations, the data for $\langle R_h \rangle$ from the experiment corresponds to the $\langle R_h \rangle$ given by the value of the line of best fit for a given chain length in Figure 2.2. Simulations

systematically slightly underestimate the experimentally determined $\langle R_h \rangle$ as shown by the slope of the line of best fit which is 0.94.

Figure 2.3 suggests that the simulations describe the experimental data well, although there is a systematic underestimation of the experimentally determined $\langle R_h \rangle$. This is not surprising for two reasons: 1) the simulations do not include the dye used to label the protein molecules in the FCS experiments; and 2) the simulations are of homopolymeric polyglutamine, whereas the experiments were performed on polyglutamine molecules of the form GQ$_N$C*KK, with C* representing a cysteine labeled with the aforementioned dye. The dye used in these experiments has a molecular weight of 720 Daltons. By comparison, Q$_{15}$ has a molecular weight of 1940 Daltons. Also, the additional amino acids (GCKK) in the experimental constructs add another 435 Daltons relative to the simulated constructs. Both the dye and additional amino acids would increase the hydrodynamic radius relative to the homopolymer. In both simulation and experiment, the exponent for the scaling of the hydrodynamic radius with chain length $\nu \approx 0.32$, indicating a chain in a poor solvent.

2.4.4 Analysis of Ratio of $\langle R_h \rangle / \langle R_g \rangle$

From the comparison of the FCS data to simulation, we conclude that the simulations appear to recapitulate the major features of the experimental system. Using the simulation results, we can determine both the hydrodynamic radius and the radius of gyration. Wu and coworkers have suggested that the ratio of the hydrodynamic radius to the radius of gyration of an individual chain can be used as an additional measure for

characterizing polymer size and shape. (*32*) Specifically, this ratio is ~0.66 for a polymer with random coil dimensions (i.e. a chain in a good solvent), ~1.29 for a rigid, uniform sphere, and a ratio higher than 1.29 indicates some amount of molten-like behavior at the outer shell of the polymer globule. Wu and coworkers found that a single poly (*N*-isopropylacrylamide) (PNIPAM) chain in water undergoes a coil-to-globule transition upon heating from 29° C to 35° C, with $\langle R_h \rangle / \langle R_g \rangle \approx 0.66$ in the coil state and $\langle R_h \rangle / \langle R_g \rangle \approx$ 1.30 in the globule state. Interestingly, there was a discontinuity in the transition between the coil and globule state where the ratio $\langle R_h \rangle / \langle R_g \rangle$ spiked to a value of 1.61. This was attributed to an additional state called the molten globule state where the core of this globule has a density similar to the fully collapsed globule, but the corona of this globule has a slightly lower density. This is illustrated in Figure 2.4, with A representing the maximally collapsed globule, B representing the molten globule, and C representing the random coil.
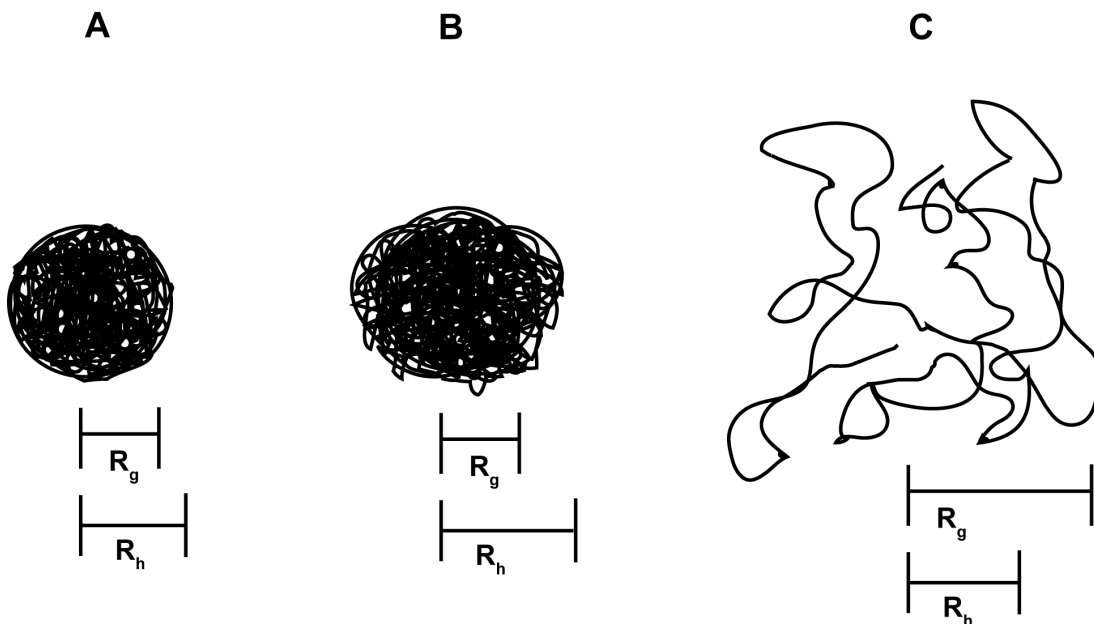
Figure 2.4 An illustration of the classes of structures adopted by a polymer in solution. A) A maximally collapsed globule with uniform density ($\langle R_h \rangle / \langle R_g \rangle \approx 1.29$). B) A molten globule with a core density similar to the maximally collapsed globule, but a region near the periphery of the globule with lower density. This leads to an increase in the hydrodynamic radius, but not the radius of gyration ($\langle R_h \rangle / \langle R_g \rangle > 1.29$). C) A random coil with a decreased hydrodynamic radius due to chain-draining, but a much larger radius of gyration ($\langle R_h \rangle / \langle R_g \rangle \approx 0.66$).

We calculated the hydrodynamic radius and the radius of gyration from simulations of the homopolymer chains Q_N (N = 15, 20, 25, 30, 35, and 45). These values and the ratio ($\langle R_h \rangle / \langle R_g \rangle$) are presented in Table 2.1. The ratio ($\langle R_h \rangle / \langle R_g \rangle$) indicates that these chains are almost maximally compact. As a comparison, Q30 was also simulated using a force-field with only attractive Leonard-Jones interactions, which should lead to a maximally collapsed globule. The ratio of the average radius of gyration

for Q_{30} simulated using the full Hamiltonian and the average radius of gyration for Q_{30} simulated with only Leonard-Jones interactions is 0.99 ± 0.02. Taken together, the data suggest that polyglutamine chains are very poorly solvated in water.

N	15	20	25	30	35	45
$\langle R_h \rangle$ (Å)	9.01	9.66	10.32	11.02	11.59	12.81
$\langle R_g \rangle$ (Å)	7.17	7.43	7.89	8.43	8.86	9.88
$\langle R_h \rangle / \langle R_g \rangle$	1.25	1.30	1.31	1.31	1.31	1.30

Table 2.1 The hydrodynamic radius ($\langle R_h \rangle$), radius of gyration ($\langle R_g \rangle$), and their ratio $\langle R_h \rangle / \langle R_g \rangle$ were determined from simulations of different chain lengths (N) of homopolymeric polyglutamine. A ratio of $\langle R_h \rangle / \langle R_g \rangle \approx 1.29$ indicates a maximally compact sphere. A ratio of $\langle R_h \rangle / \langle R_g \rangle$ higher than 1.29 indicates some amount of molten-like character around the periphery of the globule. A ratio of $\langle R_h \rangle / \langle R_g \rangle \approx 0.66$ indicates a chain in a random-coil state. Values are given in angstroms (Å).

Both CD and NMR data suggest that polyglutamine in PBS is intrinsically disordered. (*3-7*) Taken together with our data and results from molecular dynamics simulations (*2*) as well as more recent simulation results, the value of $\nu \approx 0.33$ is consistent with the interpretation that, in water, polyglutamine has access to an ensemble of distinct collapsed structures, and no single collapsed structure dominates the equilibrium distribution. The demonstration that polyglutamine polypeptides a) form collapsed structures in water and b) do not show evidence of abrupt changes in average

solution structure as chain lengths cross the pathological threshold region has important implications for both the thermodynamics and kinetics of aggregation.

2.5 Discussion

2.5.1 General implications of FCS results

The conclusion that aqueous environments are polymeric poor solvents for polyglutamine is striking because it is generally assumed that scaling of sizes with chain length follow theoretical predictions only if chains are sufficiently long ($N > 100$). (*11*) The implication of our result is that even relatively short polyglutamine peptides (15-53 residues) are akin to generic, flexible, linear polymers.

Our result for the value of ν contradicts expectations based on hydrophobicity scales. (*33-36*) According to these scales, glutamine, which is a polar amino acid, is hydrophilic. If we were to follow these hydrophobicity scales, the prediction would be that polyglutamine prefers relatively extended conformations (*4, 37*) because polar tracts are unlikely to favor collapsed structures in water. Accordingly, one might expect the value for ν to be either 0.5 or 0.59. The former would be based on the expectation that intra-chain interactions exactly counterbalance chain-solvent interactions and the latter would suggest that chain-solvent interactions are preferred to chain-chain interactions. Our results clearly demonstrate that both of these expectations are incorrect.

Although the free energy of hydration for primary and secondary amides is highly favorable (*38*), we find that even short polyglutamine chains (ca. *N*=15) prefer collapsed structures that minimize interactions with aqueous solvents. This may be explained as being the result of a "tug-of-war" between the self-association versus water solvating polar components within the chain. (*39*) For polyglutamine in water, self-association is favored.

2.5.2 FCS results are consistent with the phase behavior of polyglutamine

In a poor solvent, polymers either form collapsed globules or phase separate by aggregating. (*11, 12*) The former occurs in dilute solutions and the latter are realized as chain concentration is increased. This two-phase behavior is available only to polymers in poor as opposed to good or theta solvents. (*11*) Therefore, the observation that water acts as a poor solvent for polyglutamine is not surprising given its tendency to aggregate in aqueous solvents. (*4, 8, 9*)

Perutz et al. proposed that phase separation and aggregation of polyglutamine may be driven by the special hydrogen bonding characteristics of the glutamine sidechain. (*40*) Conversely, Dobson and coworkers have argued that the ability to form ordered aggregates is a generic attribute of polypeptide chains. (*41-44*) If water is a poor solvent for polar polyglutamine, it must be a poor solvent for generic polypeptides. If so, the driving force for aggregation for all polypeptides may be attributed to commonalities in the balance of chain-chain and chain-solvent interactions in water. Clearly, we have

much to learn about the details of polypeptide hydration and how these details influence coil-globule transitions (*45-47*) and phase behaviors of polypeptides, especially for sequences which are deficient in residues that are deemed to be hydrophobic (*47-49*).

2.5.3 Reconciling FCS results with CD and NMR data

The value of $\nu = 0.32 \pm 0.02$ obtained from analysis of our FCS data is similar to the value for ν calculated by Dima and Thirumalai for a collection of 403 folded, globular, monomeric proteins (*50*). They showed that for these systems, $\langle R_g \rangle$, the radius of gyration, scales with chain length as $N^{0.33}$. In light of these observations and expectations from polymer theories we interpret our FCS results to mean that polyglutamine prefers collapsed structures in aqueous solvents similar in global features to those adopted by folded proteins. However, as noted earlier, CD and NMR data suggest that there is marked heterogeneity in local conformational preferences for monomeric polyglutamine. Are the FCS and CD / NMR results compatible with each other? (*3-7*) Results from computational studies on conformational equilibria of monomeric polyglutamine peptides in water are consistent with both sets of observations. (*2, 51, 52*) Monomeric polyglutamine is shown to prefer a wide range of collapsed structures. Additionally, the ensemble is characterized by an absence of marked preference for distinct secondary structures (*2, 51, 52*). For a homopolymer such as polyglutamine, it is unlikely that there will be a strong preference for a unique compact conformation. This is because there is no unique way to partition glutamine residues in the chain between the interior and the surface of a globule. (*53, 54*) Consequently,

sterically allowed conformations that are consistent with the requirement of being compact are likely to be of equivalent stability.

2.5.4 Comparing FCS results to more recent experiments

Since the publication of our FCS results, at least three more experimental studies related to the collapse of polyglutamine chains have been published. In 2008, Singh and Lapidus examined the conformational preferences of polyglutamine chains of the form KKCQ$_n$WKK, with n = 4, 7, 10, 13, and 16. (*55*) In this work, the authors used the contact dependent quenching of the tryptophan triplet state by cysteine to probe chain dynamics. The details of this technique can be found elsewhere (*55, 56*). The authors found that the rate of cysteine-to-tryptophan contact increased with increasing n for n < 13, and then decreased. To interpret these results, the authors used a worm-like chain model with excluded volume (*57*) which can be used to determine $P(r)$, the probability distribution for the cysteine-to-tryptophan separation (r). This distribution can then be converted into a contact rate. (*58*) Using this formalism, the authors globally fit contact rate data for all chain lengths and arrive at the conclusion that polyglutamine chains are stiff, with a persistence length of 1.3 nm, and adopt canonical random-coil dimensions.

Singh and Lapidus assume that the conformational ensemble for all chain lengths is similar. Soon to be published simulation results generated by Nicholas Lyle from the Pappu lab suggest that the short chains used in the Singh and Lapidus work adopt alpha-helical structures, which lead to decreased end-to-end contact rates compared to the longer chain lengths. The chains appear to adopt α-helical geometries as a compromise

between the minimizing the electrostatic repulsion between lysines at the two ends of the chain, and the polyglutamine stretch compacting to sequester itself from solvent. Shorter chains form α-helices because these are locally compact structures. For longer chain lengths ($N > \sim 8$), simulations suggest that the chains can collapse fully into globules with the lysines still remaining far apart, on average. Interestingly, this transition between an ensemble dominated by α-helical geometries and an ensemble dominated by disordered globules occurs as the glutamine length approaches 15-16 residues, which is an accord with the Singh and Lapidus results. However, it is worth reiterating that the helix induced in short chains containing two lysines on either end is not seen in short chains of homopolymeric polyglutamine.

Realizing that a discrepancy existed between our results and the Singh and Lapidus results, Walters and Murphy revisited the polyglutamine collapse problem by measuring the average end-to-end distance of varying lengths of polyglutamine using Förster resonance energy transfer (FRET). Peptides of the form $K_2WQ_nXK_2$ were used for the FRET measurements, where $n = 8$, 12, 16, 20, and 24, and X is a dansylated lysine. The Förster radius (R_0) for the tryptophan-dansyl FRET pair is 2.1 nm. (*59-61*) The authors measured the steady-state FRET efficiency of each construct, and converted this efficiency to an average end-to-end distance. The calculated distances were surprisingly similar for all chain lengths (1.99 nm for Q8 to 2.4 nm for Q24) when measured in pH 7.4 PBSA (0.01 M buffer salts, 0.14 M NaCl, and 3 mM sodium azide). However, when the authors repeated the measurements in pH 12 buffer (0.01 M KCl and NaOH to adjust to pH 12) to neutralize the lysines, the calculated distances were

independent of chain length (~1.89 nm for all constructs). The authors interpreted their results as follows: 1) neutralization of the lysines has a much larger effect for long polyglutamine constructs, 2) the polyglutamine chains have a chain-length dependent persistence length, and 3) the polyglutamine chains undergo a transition from a good solvent to a poor solvent at ~Q_{16}.

A major potential complication of these experiments is the improper choice of FRET pairs. Multiple groups have shown that using FRET to measure short distances can lead to very large errors. (*62-65*) There are at least two major cautions from these works: 1) one must ensure that each fluorophore is freely rotating ; and 2) one must ensure that the distances being studied are between $0.8*R_0$ and $2.0*R_0$ because the commonly used formalism for converting FRET efficiencies to distances assumes that the dyes behave as point-dipoles. If either approximation is not met, this can lead to unreasonably large estimates of average separation distances.

One simple method of establishing free rotation of the fluorophores is to perform time-resolved fluorescence anisotropy or steady state fluorescence anisotropy as a function of protein size. In time-resolved fluorescence anisotropy, one would expect a decay of approximately 170 ps for a freely rotating tryptophan attached to a solvent-exposed portion of a protein. (*66*) In a steady-state experiment, the measured anisotropy of a freely rotating tryptophan would be independent of the size of the protein in which it is contained. Neither technique was used to check the free-rotation assumption in the study by Walters and Murphy. However, we have performed steady state fluorescence

anisotropy on polyglutamine constructs of the form WQ_nKK (for n = 30, 40, and 50), and the anisotropy increases from 0.041 mA (milli-Anisotropy units) for n = 30, to 0.054 mA for n = 40, and to 0.063 mA for n = 50. This suggests that the tryptophan in these constructs is not freely rotating, which one would assume would also be the case for the tryptophan in the constructs used by Walters and Murphy. This casts doubt on the reliability of the calculated distances in the work of Walters and Murphy.

The second approximation used in the work of Walters and Murphy to extract average distances is that the dyes exist as point-dipoles. Sahoo et al (*64*) suggest this approximation is only valid if the distances being probed are between $0.8*R_0$ and $2.0*R_0$, which translates to distances between 1.69 nm and 4.20 nm for the FRET pair used by Walters and Murphy. The average distances calculated by Walters and Murphy lie between 1.89 nm and 2.4 nm. The average distances, especially for the shortest chain lengths and the experiments performed at pH 12, lie close to the 1.69 nm cutoff. If the chains adopt conformations which bring the dyes closer than this, these conformations will not be appropriately represented in the calculated average distance, leading to calculated distances larger than reality.

As an example of how problematic restricted fluorophore motion and a breakdown in the point-dipole approximation can be in determining end-to-end distances, we point to the work of Schuler et al. (*67*) The authors used single-molecule FRET to investigate the length distribution of polyprolines containing 6, 11, 14, 20, 27, 33, and 40 prolines. They found that the shorter chain lengths (<20) had significantly lower FRET

efficiencies, and, hence, significantly higher calculated distances (~3.3 nm for P6), than expected (~1.7 nm for P6) based on even the longest possible end-to-end distance of these polyprolines. An analysis of the time-resolved anisotropy of the donor dye showed that the dye was not freely rotating. In addition, the FRET pair they were using had an R_0 of 5.4 nm, which would only be useful probing distances higher than ~4.3 nm according to the $0.8*R_0$ limit. A subsequent experiment by Sahoo et al. using a FRET pair with an R_0 of 0.9 nm showed polyproline 6 to have an average end-to-end distance of only 1.66 nm. (*64*) This is less than half the distance calculated by Schuler and colleagues, suggesting the calculated end-to-end distances are extremely sensitive to the FRET pair's R_0 and the free rotation of the fluorophores.

Simulations performed by Lyle in the Pappu lab[3] suggest that the average cysteine-to-tryptophan distance for constructs of the form $K_2CQ_nWK_2$ is only 0.9 nm for $n = 7$ and 1.3 nm for $n = 24$. If true, the experiments used by Walters and Murphy would be unable to reliably extract these distances even if the fluorophore motion was unrestricted.

Although the foregoing caveats make the work of Walters and Murphy difficult to interpret quantitatively, it is worth noting that the trends they observe are likely reliable. These trends a) are consistent with the collapse of polyglutamine in aqueous solution; b) are recapitulated by Lyle's unpublished results; c) disagree with a polyglutamine length-

[3] Lyle, N., Crick, S.L., and Pappu, R.V. *Manuscript in Preparation*

dependent change in solvent quality; and d) help explain the disconnect between Lapidus and Crick et al.

Dougan and coworkers used force-clamp spectroscopy with atomic force microscopy (AFM) to measure the collapse of polyglutamine chains and the mechanical stability of these globules. (*68*) Polyglutamine insertions (*n*=15, 25, 50 and 75) were placed into a polyprotein construct of the form I27-Qn-I27-Qn-I27, where I27 represents the human cardiac titin Ig domain I27. The I27 domain has been well characterized mechanically. (*69-71*)   The unfolding of I27 by force gives an easily identifiable mechanical fingerprint, which allows the Q$_n$ segments to be probed independently. When the polyprotein construct containing polyglutamine was stretched under a constant force of 180 pN using AFM, the construct unfolded in step increases of 24 nm, which is what is expected for the I27 domains alone. Moreover, measurements of the initial lengths of the constructs containing polyglutamine are essentially indistinguishable from the constructs not containing the polyglutamine.   Both of these results hold independent of the polyglutamine chain length suggesting that polyglutamine molecules are not only collapsed, but are more mechanically stable than folded I27 domains. This result adds support to our conclusions.

2.6 Summary

Using FCS we have quantified the scaling of hydrodynamic size as a function of chain length for monomeric polyglutamine and demonstrated that water acts a poor solvent for polyglutamine. Attractions between chain residues are preferred to interactions with the surrounding solvent. Therefore, chains either collapse to form globular structures or, as concentration increases, they form intermolecular aggregates. (*11, 12*) Our observation that water is a poor solvent for monomeric polyglutamine implies that there is a generic driving force for polyglutamine aggregation. Additionally, if monomeric polyglutamine does not show a clear preference for a specific globular conformation (a proposal supported by CD and NMR data and results from molecular dynamics simulations), at equilibrium a heterogeneous ensemble of globular conformations are preferred. This preference for conformational heterogeneity (*i.e.* disorder) provides an additional driving force for folding via aggregation.

Although structural and thermodynamic studies place necessary constraints on models for nucleation of aggregation, they cannot be used to adjudicate between different mechanisms. It is important to reiterate that a range of nucleation mechanisms are possible, especially if one views aggregation as being analogous to polymer aggregation / crystallization. Evidence of different mechanisms can be inferred from the presence of on or off-pathway intermediates (*72-74*), the absence of intermediates under certain conditions (*4-6, 10, 21, 75*), variations in morphologies with solution conditions (*75*), and the fact that nucleus size can depend on protein concentration (*76, 77*). The dominant

mechanism of aggregation will vary with solution conditions and it is important to

uncover the family of possible mechanisms rather than focusing on any one mechanism.

2.7 References

1. Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R., and Pappu, R. V. (2006) Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions, *Proc Natl Acad Sci U S A 103*, 16764-16769.
2. Wang, X. L., Vitalis, A., Wyczalkowski, M. A., and Pappu, R. V. (2006) Characterizing the conformational ensemble of monomeric polyglutamine, *Proteins-Structure Function And Bioinformatics 63*, 297-311.
3. Altschuler, E. L., Hud, N. V., Mazrimas, J. A., and Rupp, B. (1997) Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides, *Journal Of Peptide Research 50*, 73-75.
4. Wetzel, R. (2005) In *Protein Folding Handbook* (Buchner, J., and Kiefhaber, T., Eds.), Wiley-VSH Verlag GmbH & Co Weinheim.
5. Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001) Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity, *Journal Of Molecular Biology 311*, 173-182.
6. Chen, S., Ferrone, F. A., and Wetzel, R. (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation., *Proc Natl Acad Sci U S A 99*, 11884-11889.
7. Masino, L., Kelly, G., Leonard, K., Trottier, Y., and Pastore, A. (2002) Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins, *FEBS Letters 513*, 267-272.
8. Krull, L. H., Wall, J. S., Zobel, H., and Dimler, R. J. (1965) Synthetic Polypeptides Containing Side-Chain Amide Groups: Water-insoluble Polymers*, *Biochemistry 4*, 626-633.
9. Krull, L. H., and Wall, J. S. (1966) Synthetic Polypeptides Containing Side-Chain Amide Groups. Water-Soluble Polymers*, *Biochemistry 5*, 1521-1527.
10. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction., *Proc Natl Acad Sci U S A 102*, 15400-15405.
11. Rubinstein, M., and Colby, R. H. (2003) *Polymer Physics*, Oxford University Press, Oxford and New York.
12. Chan, H. S., and Dill, K. A. (1991) Polymer principles in protein structure and stability, *Annu Rev Biophys Biophys Chem 20*, 447-490.

13. Tran, H. T., and Pappu, R. V. (2006) Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions, *Biophysical Journal 91*, 1868-1886.

14. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S., and Plaxco, K. W. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins, *Proc Natl Acad Sci U S A 101*, 12491-12496.

15. Flory, P. J. (1953) *Principles of Polymer Chemistry*, Cornell University Press, Ithaca and London.

16. Flory, P. J. (1969) In *Statistical Mechanics of Chain Molecules*, pp 1-48, Hanser Publishers, New York.

17. Elson, E. L., and Magde, D. (1974) Fluorescence correlation spectroscopy. I. Conceptual basis and theory, *Biopolymers 13*, 1-27.

18. Rigler, R., and Elson, E. S. (2001) *Fluorescence Correlation Spectroscopy*, Springer, Verlag, Berlin, Heidelberg.

19. Rigler, R., Mets, Ü., Widengren, J., and Kask, P. (1993) Fluorescence correlation spectroscopy with high count rate and low background: analysis of translational diffusion, *European Biophysics Journal 22*, 169-175.

20. Magde, D., Elson, E. L., and Webb, W. W. (1974) Fluorescence correlation spectroscopy. II. An experimental realization, *Biopolymers 13*, 29-61.

21. Chen, S., and Wetzel, R. (2001) Solubilization and disaggregation of polyglutamine peptides, *Protein Sci 10*, 887-891.

22. Gentle, J. E. (2002) In *Elements of computational statistics* pp 69-98, Springer-Verlag, New York.

23. Pristinski, D., Kozlovskaya, V., and Sukhishvili, S. A. (2005) Fluorescence correlation spectroscopy studies of diffusion of a weak polyelectrolyte in aqueous solutions, *J Chem Phys 122*, 14907.

24. Petrasek, Z., and Schwille, P. (2008) Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy, *Biophys J 94*, 1437-1448.

25. Vishwanath, D. S., and Natarajan, G. (1989) *Data book on the viscosity of liquids* Hemisphere Publishing Corp, New York.

26. Vitalis, A., and Pappu, R. V. (2009) ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions, *J Comput Chem 30*, 673-699.

27. Vitalis, A., and Pappu, R. V. (2009) Methods for Monte Carlo simulations of biomacromolecules, *Annu Rep Comput Chem 5*, 49-76.

28. Williamson, T. E., Vitalis, A., Crick, S. L., and Pappu, R. V. (2010) Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin, *J Mol Biol 396*, 1295-1309.

29. Garcia De La Torre, J., Huertas, M. L., and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure, *Biophys J 78*, 719-730.

30. Carrasco, B., Garcia de la Torre, J., and Zipper, P. (1999) Calculation of hydrodynamic properties of macromolecular bead models with overlapping spheres, *Eur Biophys J 28*, 510-515.

31. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., and Pappu, R. V. (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins, *Proc Natl Acad Sci U S A 107*, 8183-8188.

32. Wu, C., and Zhou, S. (1996) First Observation of the Molten Globule State of a Single Homopolymer Chain, *Phys Rev Lett 77*, 3053-3055.

33. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985) Hydrophobicity of Amino-Acid Residues in Globular-Proteins, *Science 229*, 834-838.

34. Kyte, J., and Doolittle, R. F. (1982) A Simple Method for Displaying the Hydropathic Character of a Protein, *Journal of Molecular Biology 157*, 105-132.

35. Janin, J. (1979) Surface and inside Volumes in Globular Proteins, *Nature 277*, 491-492.

36. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot, *Journal of Molecular Biology 179*, 125-142.

37. Uversky, V. N., and Fink, A. L. (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded, *Biochim Biophys Acta 1698*, 131-153.

38. Wolfenden, R. (1978) Interaction of the peptide bond with solvent water: a vapor phase analysis, *Biochemistry 17*, 201-204.

39. Yalkowski, S. H. (1999) In *Solubility and solubilization in aqueous media*, pp 39-45, Oxford University Press, Oxford.

40. Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994) Glutamine Repeats as Polar Zippers - Their Possible Role in Inherited Neurodegenerative Diseases, *P Natl Acad Sci USA 91*, 5355-5358.

41. Dobson, C. M. (2004) Principles of protein folding, misfolding and aggregation, *Semin Cell Dev Biol 15*, 3-16.

42. Fandrich, M., Fletcher, M. A., and Dobson, C. M. (2001) Amyloid fibrils from muscle myoglobin - Even an ordinary globular protein can assume a rogue guise if conditions are right., *Nature 410*, 165-166.

43. Fandrich, M., and Dobson, C. M. (2002) The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation, *Embo J 21*, 5682-5690.

44. Dobson, C. M. (1999) Protein misfolding, evolution and disease, *Trends Biochem Sci 24*, 329-332.

45. Zagrovic, B., Lipfert, J., Sorin, E. J., Millettt, I. S., van Gunsteren, W. F., Doniach, S., and Pande, V. S. (2005) Unusual compactness of a polyproline type II structure, *P Natl Acad Sci USA 102*, 11698-11703.

46. Sherman, E., and Haran, G. (2006) Coil-globule transition in the denatured state of a small protein, *P Natl Acad Sci USA 103*, 11539-11543.

47. Moglich, A., Joder, K., and Kiefhaber, T. (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate

intramolecular hydrogen bond formation, *P Natl Acad Sci USA 103*, 12394-12399.

48. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci 11*, 739-756.

49. Dunker, A. K., Brown, C. J., and Obradovic, Z. (2002) Identification and functions of usefully disordered proteins, *Adv Protein Chem 62*, 25-49.

50. Dima, R. I., and Thirumalai, D. (2004) Asymmetry in the shapes of folded and denatured states of proteins, *J Phys Chem B 108*, 6564-6570.

51. Vitalis, A., Wang, X., and Pappu, R. V. (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization., *J Mol Biol 384*, 279-297.

52. Vitalis, A., Wang, X., and Pappu, R. V. (2007) Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories., *Biophys J 93*, 1923-1937.

53. Dill, K. A. (1990) Dominant Forces in Protein Folding, *Biochemistry 29*, 7133-7155.

54. Dill, K. A. (1999) Polymer principles and protein folding, *Protein Science 8*, 1166-1180.

55. Singh, V. R., and Lapidus, L. J. (2008) The intrinsic stiffness of polyglutamine peptides, *J Phys Chem B 112*, 13172-13176.

56. Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2000) Measuring the rate of intramolecular contact formation in polypeptides, *Proc Natl Acad Sci U S A 97*, 7220-7225.

57. Szabo, A., Schulten, K., and Schulten, Z. (1980) *J. Chem. Phys. 72*, 4350.

58. Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2001) *Phys. Rev. Lett. 87*, 258101.

59. Gustiananda, M., Liggins, J. R., Cummins, P. L., and Gready, J. E. (2004) Conformation of prion protein repeat peptides probed by FRET measurements and molecular dynamics simulations, *Biophys J 86*, 2467-2483.

60. Moglich, A., Joder, K., and Kiefhaber, T. (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation, *Proc Natl Acad Sci U S A 103*, 12394-12399.

61. Soranno, A., Longhi, R., Bellini, T., and Buscaglia, M. (2009) Kinetics of contact formation and end-to-end distance distributions of swollen disordered peptides, *Biophys J 96*, 1515-1528.

62. Speelman, A. L., Muñoz-Losa, A., Hinkle, K. L., VanBeek, D. B., Mennucci, B., and Krueger, B. P. (2011) Using Molecular Dynamics and Quantum Mechanics Calculations To Model Fluorescence Observables, *The Journal of Physical Chemistry A 115*, 3997-4008.

63. Sahoo, H., Roccatano, D., Zacharias, M., and Nau, W. M. (2006) Distance Distributions of Short Polypeptides Recovered by Fluorescence Resonance Energy Transfer in the 10 Å Domain, *Journal of the American Chemical Society 128*, 8118-8119.

64. Sahoo, H., Roccatano, D., Hennig, A., and Nau, W. M. (2007) A 10-Å Spectroscopic Ruler Applied to Short Polyprolines, *Journal of the American Chemical Society 129*, 9762-9772.

65. Dolghih, E., Ortiz, W., Kim, S., Krueger, B. P., Krause, J. L., and Roitberg, A. E. (2009) Theoretical Studies of Short Polyproline Systems: Recalibration of a Molecular Ruler†, *The Journal of Physical Chemistry A 113*, 4639-4646.

66. Lakowicz, J. R. (2006) *Principles of fluorescence spectroscopy*, 3rd ed., Springer, New York.

67. Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M., and Eaton, W. A. (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence, *Proc Natl Acad Sci U S A 102*, 2754-2759.

68. Dougan, L., Li, J., Badilla, C. L., Berne, B. J., and Fernandez, J. M. (2009) Single homopolypeptide chains collapse into mechanically rigid conformations, *Proc Natl Acad Sci U S A 106*, 12605-12610.

69. Li, H., Linke, W. A., Oberhauser, A. F., Carrion-Vazquez, M., Kerkvliet, J. G., Lu, H., Marszalek, P. E., and Fernandez, J. M. (2002) Reverse engineering of the giant muscle protein titin, *Nature 418*, 998-1002.

70. Dougan, L., Feng, G., Lu, H., and Fernandez, J. M. (2008) Solvent molecules bridge the mechanical unfolding transition state of a protein, *Proc Natl Acad Sci U S A 105*, 3185-3190.

71. Garcia-Manyes, S., Brujic, J., Badilla, C. L., and Fernandez, J. M. (2007) Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin, *Biophys J 93*, 2436-2446.

72. Wacker, J. L., Zareie, M. H., Fong, H., Sarikaya, M., and Muchowski, P. J. (2004) Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer, *Nat Struct Mol Biol 11*, 1215-1222.

73. Ross, C. A., and Poirier, M. A. (2005) Opinion: What is the role of protein aggregation in neurodegeneration?, *Nat Rev Mol Cell Biol 6*, 891-898.

74. Gosal, W. S., Morten, I. J., Hewitt, E. W., Smith, D. A., Thomson, N. H., and Radford, S. E. (2005) Competing pathways determine fibril morphology in the self-assembly of beta2-microglobulin into amyloid, *J Mol Biol 351*, 850-864.

75. Chen, S. M., Berthelier, V., Hamilton, J. B., O'Nuallain, B., and Wetzel, R. (2002) Amyloid-like features of polyglutamine aggregates and their assembly kinetics, *Biochemistry 41*, 7391-7399.

76. Powers, E. T., and Powers, D. L. (2006) The kinetics of nucleated polymerizations at high concentrations: amyloid fibril formation near and above the "supercritical concentration", *Biophys J 91*, 122-132.

77. Frieden, C., and Goddette, D. W. (1983) Polymerization of actin and actin-like systems: evaluation of the time course of polymerization in relation to the mechanism, *Biochemistry 22*, 5836-5843.

Chapter 3: Constructing a Phase Diagram for Polyglutamine and the Phase Diagram's Implications[1]

3.1 Preamble

We have shown that aqueous solutions at 25° C represent poor solvent conditions for polyglutamine. In dilute solutions, polyglutamine molecules form a dispersed solution of globules. As the concentration increases, the globules will coalesce to form clusters resulting in demixing and eventual phase separation. The details of this process require knowledge of the phase behavior of polyglutamine solutions. In this chapter, we describe phase diagrams for aqueous polyglutamine solutions and discuss the implications of this phase diagram on possible mechanisms of phase separation.

[1]The material in this chapter is adapted from:
S.L.Crick, R.V. Pappu. (2010).Thermodynamic and Kinetic Models for Aggregation of Intrinsically Disordered Proteins. Peptide Folding, Misfolding, and Unfolding. Ed. Reinhard Schweitzer-Stenner. John Wiley & Sons, Hoboken, NJ. In press.

and

S.L. Crick, A.H. Mao, C. Frieden, R.V. Pappu, Inferring aggregation mechanisms for polyglutamine peptides from phase diagrams, (2011) In preparation.

3.2 Introduction

3.2.1 Overview

Like most generic carbon-based polymers, proteins are only marginally soluble in aqueous solvents. Experiments have shown that protein solubility for folded proteins decreases as the stability of the folded state decreases. (*1*) This is likely due to increased solvent exposure of hydrophobic groups which folded proteins tend to partition in the core of the protein. (*2*) It has long been thought that the "hydrophobic effect" (*3*) is responsible for much of the organization, such as protein folding, that takes place in biological systems. (*4*) Thus, partitioning of hydrophobic groups away from solvent via protein folding may be viewed as a strategy to keep proteins soluble. Moreover, it appears that there is a constant competition between the formation of a protein's native fold and association into aggregates, and this is likely kinetically controlled. (*5-7*) Hydrophobic groups tend to associate (*8*) to sequester themselves from water. At higher protein concentrations, this will lead to protein aggregation. At lower concentrations, the fold can be achieved before aggregation occurs. Both cases allow the protein-solvent system to minimize its free energy, and a major component of this free energy minimization is the exclusion of water from hydrophobic regions of the protein. (*9-11*)

Although polyglutamine is comprised of amino acids not typically considered hydrophobic, polyglutamine prefers to associate with itself in aqueous solutions as opposed to forming extensive interactions with water. (*12*) Polyglutamine does not

satisfy the relatively narrow IUPAC definition of a hydrophobe (*13*), but it does self-solvate and preferentially exclude water, which are common features of hydrophobic groups in proteins. (*4, 14*)

If polyglutamine is comprised of polar amino acids, why does it behave like a hydrophobic molecule? It has been shown for both synthetic polymers (*15*) and proteins (*16*) that intrachain hydrogen bonds (of which polyglutamine is capable of making many) lead to this non-classical hydrophobic effect.

In very dilute solutions like those used for the experiments in Chapter 2, polyglutamine molecules are homogeneously distributed in solution and exist as compact globules. Although globule formation is the best way for an isolated monomer to exclude the most glutamines from solvent, this still leaves glutamines on the surface of globules. The number of glutamines on the surface scales as $N^{2/3}$, where N is chain length. As the concentration of polyglutamine in solution increases, the drive to sequester the glutamines from the solvent leads to phase separation.

Phase separation of polyglutamine shows the following characteristics: 1) there is a protein concentration below which phase separation will not occur, 2) when phase separation does occur, it leads to the formation of insoluble, protein rich material, and 3) the time scale of phase separation depends on the protein concentration in bulk. (*17-19*) These characteristics are consistent with phase separation of polymer and colloidal

solutions. (*20*) This should not be surprising since polyglutamine, like all proteins, is a

polymer, but can also be considered a colloid (*21*). These features allow us to touch base

with the rich theoretical framework of polymer physics (*22-28*) and colloidal science (*29-

32*) in the study of the phase separation of polyglutamine solutions.

Polymer physics and colloidal science have been applied to the study of phase

separation of proteins (*21, 33-36*), but, oddly, only in the context of protein

crystallization of folded proteins. To our knowledge, there is no known application of

these principles to the study of phase-separation/aggregation of disease related proteins,

polyglutamine included. However, these proteins are the ones for which the phase

separation seems to be a major problem, and, thus, the principles from polymer and

colloidal science would be most applicable. We use these concepts for the study of the

phase separation of polyglutamine because they provide the optimal way of connecting

theory, simulation, and experimental data.

In this chapter, we describe and implement a methodology for constructing a

phase diagram for polyglutamine solutions. We first measure the saturation curves for

solutions of $K_2Q_{30}K_2$, $K_2Q_{40}K_2$, $Q_{30}K_2$, and $Q_{40}K_2$. We then use theories from polymer

physics to construct the rest of the phase diagram. Finally, we interpret these phase

diagrams using theories from polymer physics and colloidal science.

The questions we aim to answer are as follows: How does the length of the polyglutamine segment affect the saturation curve? Do flanking sequences such as lysines used in synthetic peptides affect the saturation curves, and, hence, the thermodynamics of phase separation? It has been suggested that the lysines only affect the kinetics of the phase separation process and not the thermodynamics, but this has never been tested. (*17, 18*) Can we use information contained in phase diagrams to make inferences regarding mechanism of phase separation?

3.3 Methods

3.3.1 Preparation of Peptide Samples

All peptides were purchased in crude form from Yale University's Keck Biotechnology center. The peptides were synthesized using solid-phase synthesis according to the general designs $K_2Q_NK_2$ and Q_NK_2 for ($N = 30$ and 40) where N denotes the length of the polyglutamine segment. The lysine residues were included to increase peptide solubility. The crude material was disaggregated (*37, 38*) and purified using reverse-phase HPLC on a preparative C_3 reverse-phase column (Agilent). Peptides were stored in lyophilized form at -20° C until use.

3.3.2 Saturation-curve determination

The experimental procedure for determination of the saturation curve (C_{sat}(T)) of a given peptide construct is illustrated in Figures 3.3 A and B.

Immediately prior to performing an experiment, an appropriate amount of peptide was weighed and dissolved in a Wheaton glass vial containing a 1 mg : 1 mL : 1 mL mixture of peptide : TFA : HFIP. This solution was sonicated for 30 seconds and left to incubate at room temperature for 1 hour. Next, the solution was evaporated under a gentle nitrogen stream leaving behind a clear peptide film on the walls of the glass vial. The peptides were resuspended from this film in an appropriate amount of ultrapure water at room temperature while sonicating. The solution was then buffered to pH 7.0 using 500 mM phosphate buffer stock diluted to a final concentration of 50 mM. The solution was aliquoted (1.5 mL) into 1.6 mL Eppendorf tubes which were capped and sealed with Parafilm.

The samples were incubated in water baths at a specified temperature for two weeks. Temperatures below room temperature were maintained in water baths placed in a 4° C cold room. For each temperature, peptides were incubated at a concentration of 50 μM and 100 μM. The incubation at two different concentrations provided confidence that the phase separation was complete if both samples contained the same concentration of material in the supernatant at the end of two weeks. If, for a given temperature and

peptide construct, no phase separation was evident from visual inspection and supernatant concentration measurement after the two week incubation, the experiment was repeated with a starting peptide concentration of 1 mM.

Peptide concentrations were determined using the Micro BCA (Bicinchoninic Acid) assay (Pierce, Rockford, IL). The BCA assay (*39*) is an analog of the classical biuret-reaction (*40*) whereby one cupric ion (Cu^{2+}) forms a coordination complex with four to six nearby peptides bonds which results in the cupric ion being reduced. This leads to a mild color change at 562 nm, and the intensity of the color produced is proportional to the number of peptide bonds in solution. The BCA assay enhances the color production, and, hence the sensitivity, by the addition of bicinchoninic acid to the reaction. Two BCA molecules chelate with each Cu^{1+} ion produced and this enhances the color change at 562 nm. The Micro BCA assay is sensitive to protein concentrations as low as 0.0005 mg/mL. This corresponds to a concentration of ~70 nM for a 5 kDa protein.

The assay was calibrated using the peptide ($WQ_{40}KK$) as a standard. This peptide was disaggregated using the aforementioned protocol and resuspended in 50 mM phosphate buffer (pH 7.0). The concentration of this solution was determined using the peptide's molar extinction coefficient (5502 $M^{-1}cm^{-1}$) at 280 nm. A calibration curve for the Micro BCA assay was constructed by measuring the color development at 562 nm in solutions covering the range of concentrations from 50 nM to 10 µM after incubation

with the Micro BCA reagents at 60° C for 1 hour. To measure unknown concentrations of other polyglutamine peptides, an aliquot of the solution was removed and incubated with the Micro BCA reagents at 60° C for 1 hour after which time the color development at 562 nm was compared to the calibration curve.

After two weeks of incubation, the samples were centrifuged for 100 minutes at 25,000×G. Two aliquots of the supernatant were removed and the Micro BCA assay was performed. These aliquots had a volume ratio of 1:2 because this should lead to a color development of 1:2 in the Micro BCA assay. A 1:2 ratio in the color development of the samples provided confidence in the measurement and ensured that the measured concentrations were in the linear region of the calibration curve.
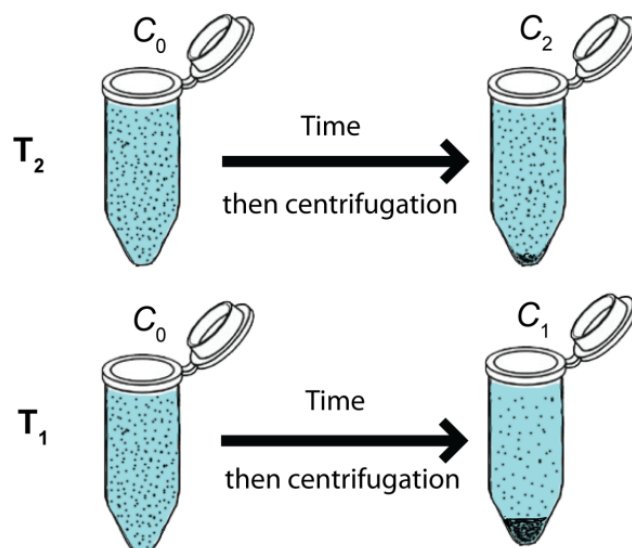
Figure 3.1 A) This diagram illustrates the procedure for determining the saturation curves of polyglutamine solutions. In this illustration, two aqueous (blue) polyglutamine (black dots) solutions at a starting concentration (C_0) are incubated at two different temperatures (T_1 and T_2 where $T_2 > T_1$) in Eppendorf tubes. With time, the solutions phase separate, and, after centrifugation, a protein-rich phase is found at the bottom of the Eppendorf tube with a solvent rich phase found in the supernatant. As shown in Figure 3.1B), the concentration of the protein in the supernatant corresponds to the saturation curve at a given temperature ((C_1,T_1) and (C_2,T_2)). This illustration and the one in Figure 3.1B correspond to a portion of a phase diagram for a system exhibiting an upper critical solution temperature (UCST).
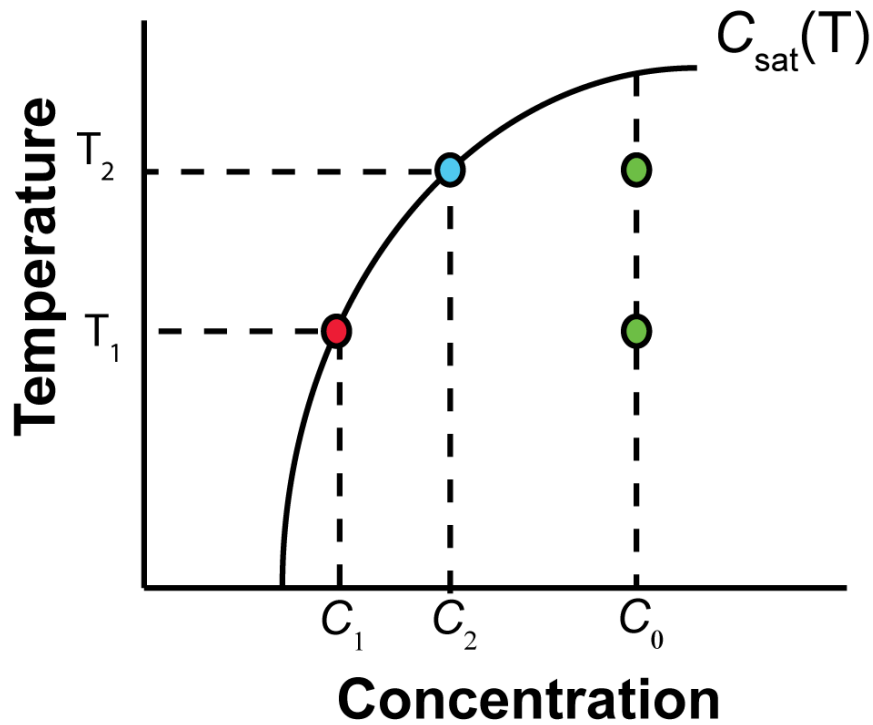
Figure 3.1 B) This diagram shows how the experiment described in Figure 3.1 A) corresponds to the determination of the saturation curve (C_{sat}(T)). Two solutions (green dots) are incubated at the same starting concentration (C_0) at two different temperatures (T_1 and T_2). After time and centrifugation (as described in Figure 3.1A), the concentrations remaining in the supernatant are given by C_1 for T_1 (red dot) and C_2 for T_2 (blue dot). These points help define the saturation curve. Repeating this process for multiple temperatures leads to the measurement of the full saturation curve.

3.3.3 Construction of the Phase Diagram

We use classical Flory-Huggins polymer theory (*41, 42*) to fit our saturation curve data and predict the remainder of the phase diagram. It is worth emphasizing that this theory has been traditionally reserved for describing liquid-liquid demixing, although there is precedent for its use in describing the precipitation from a liquid solution. Its weakness arises in distinguishing different solid phases, because it is a theory that focuses on the compositions of distinct phases and does not follow additional order parameters such as density, correlation lengths, or the conformational features of polymers. These generalizations are now available thanks to the work of Muthukumar (*43*), but they need measurements that go beyond assessment of saturation curves. Prior to describing the procedure, it is useful to provide an overview of the thermodynamics of phase separation.

3.3.3.1 Thermodynamics of Phase Separation

Aqueous solutions of polyglutamine peptides can be thought of as two-component systems comprising of the buffer solvent and the peptide at a particular concentration (in molar or molal units) or volume fraction (ϕ). For a two-component system, one should expect two limiting behaviors for the composition of elemental volumes within a peptide + solvent system. In a good solvent, the system forms a uniform mixture irrespective of the bulk concentration of the peptide / polymer. Under such circumstances, an average elemental volume within the system will have a polymer and solvent composition (concentration) that is congruent with the bulk concentrations of these entities *viz*., ϕ and $(1-\phi)$, respectively. The opposite scenario prevails in a poor

111

solvent where the system can form at least two if not more distinct phases. For the peptide + solvent system of interest, there will be at least one concentration interval $\phi' \leq \phi \leq \phi''$ for which the system partitions into a solvent-rich and a polymer-rich phase. Within the former, the peptides / polymers are dispersed in the solvent and the concentration of polymer in an average elemental volume is lower than the bulk concentration. Conversely, in the polymer-rich phase, molecules of the solvent are dispersed in a predominantly polymer matrix and the concentration of polymer in an average elemental volume within this phase is higher than the bulk concentration.

Mixtures are either homogeneous (as is the case in a good solvent where only the well-mixed phase prevails for all values of ϕ) or heterogeneous (as is the case for polymers in poor solvents where the mixture consists of least two and maybe more phases depending on the composition ϕ and thermodynamic parameters such as temperature and pressure). The driving force for forming homogeneous mixture can be quantified in terms of ΔG_{mix}, which is the free energy of mixing. The Flory-Huggins formalism (*23, 28, 41*) provides a simple albeit robust approach for quantitative characterization of ΔG_{mix} as a function of ϕ for binary, ternary, and quaternary mixtures. This is a mean-field theory that uses the device of a finite lattice to help with the enumeration of feasible arrangements of solvent molecules around the "average" conformation of a polymer. Although there are several weaknesses inherent to the random-mixing formalism of the Flory-Huggins model, it is surprisingly useful for

providing semi-quantitative descriptions of polymer solutions. According to this theory, the free energy of mixing for a two-component system is written as:

$$\frac{\Delta \overline{G}_{mix}}{RT} = \left[\frac{\phi}{n} \ln \phi + (1-\phi) \ln (1-\phi) + \chi \phi (1-\phi) \right] = \frac{\Delta \overline{H}_{mix} - T\Delta \overline{S}_{mix}}{RT}$$

[1]

$$\chi = \frac{z}{2} \frac{(2h_{PS} - h_{PP} - h_{SS})}{RT}$$

where χ is the Flory interaction parameter. It reports on the balance between per-site enthalpies of polymer-solvent (h_{PS}), polymer-polymer (h_{PP}), and solvent-solvent (h_{SS}) interactions. (23, 28, 41)

In equation [1], $\Delta \overline{G}_{mix}$ is the free energy of mixing per lattice site ($\Delta \overline{H}_{mix}$ and $\Delta \overline{S}_{mix}$ are the corresponding contributions from enthalpy and entropy to the free energy of mixing), n is the number of lattice sites occupied by the polymer on a lattice with coordination number z, $R=1.987\times10^{-3}$ kcal/mol is the gas constant, T is the system temperature, ϕ is the volume fraction of polymer, and $(1-\phi)$ is the volume fraction of solvent.

In the Flory-Huggins mean-field theory, entropy always favors mixing as shown in Figure 3.2. In a good solvent, there is a net attraction between the polymer and solvent sites and $\chi < 0$ because $2|h_{PS}| > (|h_{PP} - h_{SS}|)$. An ideal mixture corresponds to the case of $\chi = 0$ implying that $2|h_{PS}|=(|h_{PP} - h_{SS}|)$. In a poor solvent, there is net repulsion between unlike species, and $\chi > 0$. Under these conditions, the local concavity / convexity

113

determines the stability of the mixed (homogeneous) phase vis-à-vis the phase separated state.
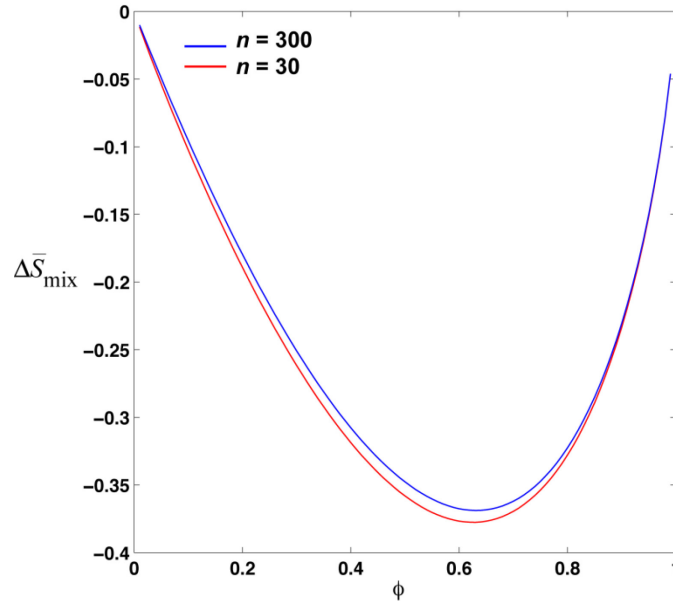
Figure 3.2: The Flory-Huggins entropy of mixing per lattice site for the case of a binary mixture. The function is convex for all values of ϕ and is indicative of entropy favoring mixing in the random mixture model. As n increases, there is a diminution in the favorability of the entropy of mixing.

When $\chi > 0$, $\Delta \bar{G}_{mix}$ plotted against ϕ will have regions of ϕ for which $\Delta \bar{G}_{mix}$ is locally concave. For these values of ϕ, the mixed state is unstable and the system separates into a solvent-rich and polymer-rich phase. The compositions of these phases are determined using the so-called common tangent rule as illustrated in Figure 3.3. This

method to determine ϕ' and ϕ'' reflects the fact that for all compositions in the interval ϕ' $\leq \phi \leq \phi''$, the system separates into the two phases (solvent-rich and polymer-rich) and the compositions of these phases namely, ϕ' and ϕ'' are determined by requiring that the chemical potentials of the coexisting solvent-rich and polymer-rich phases be equal at equilibrium ($\mu' = \mu''$). The common tangent rule requires that:

$$[2] \qquad \left.\frac{\partial \Delta \overline{G}_{\text{mix}}}{\partial \phi}\right|_{\phi=\phi'} = \left.\frac{\partial \Delta \overline{G}_{\text{mix}}}{\partial \phi}\right|_{\phi=\phi''}$$

and this is consistent with the condition for phase equilibria. The equilibrium compositions ϕ' and ϕ'' will change with temperature and the envelope of different ϕ' and ϕ'' values at different temperatures yields the so-called binodal, which is equivalent to the coexistence curve for a two-phase, two-component system. The stability of the mixed versus demixed phase is determined by the second derivative of $\Delta \overline{G}_{\text{mix}}$ such that

$\dfrac{\partial^2 \Delta \overline{G}_{\text{mix}}}{\partial \phi^2} < 0$ when the mixed state is unstable and $\dfrac{\partial^2 \Delta \overline{G}_{\text{mix}}}{\partial \phi^2} > 0$ when the mixed state is

locally stable i.e., metastable. An envelope of points can be generated in a T-ϕ plane where each point on the curve is the value of ϕ corresponding to an inflection point of

$\Delta \overline{G}_{\text{mix}}$ satisfying the condition $\dfrac{\partial^2 \Delta \overline{G}_{\text{mix}}}{\partial \phi^2} = 0$. This envelope of points is referred to as the

spinodal. The mixed state is metastable for values of ϕ that lie between the binodal and the spinodal. Within this region, achievement of the thermodynamically favored phase separated state is barrier limited and requires the formation of a critical nucleus that can

grow. The larger the gap between the binodal and spinodal, the larger the metastability interval, which in turn implies a greater likelihood of sampling metastable phases or so-called mesophases that have composition profiles that are distinct from the two equilibrium phases.
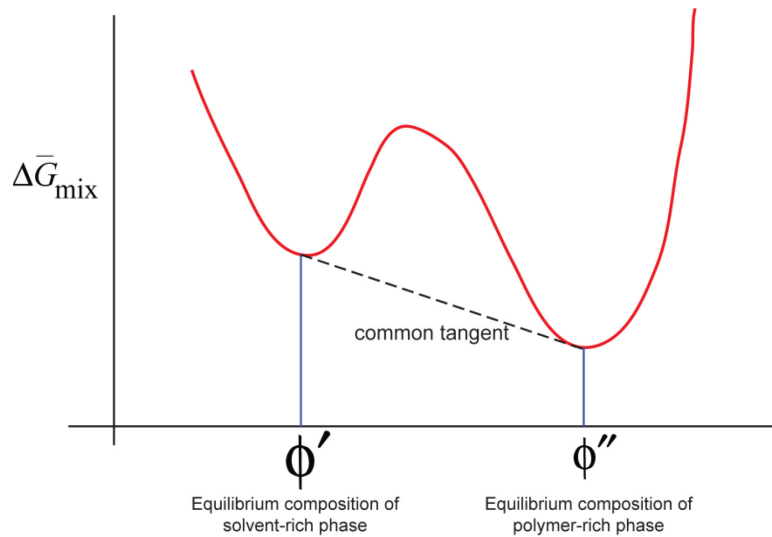
Figure 3.3 Composition dependence of the free energy of mixing for $\chi > 0$ at a temperature where phase separation is thermodynamically favorable for values of ϕ in the interval $\phi' \leq \phi \leq \phi''$. The values of ϕ' and ϕ'' at a given temperature T are inferred using the common tangent rule depicted using the dashed line.

3.3.3.2 Procedure for Constructing the Phase Diagram

In our system, ϕ and $(1 - \phi)$ represent the volume fractions of polymer and solvent, respectively. Because the theory is derived from a lattice model, n is supposed to represent the number of lattice sites occupied by the polymer. We set n equal to the ratio of the number of water molecules excluded by a polymer chain assuming the chain to be a globule with a mass density of 1.43 g/cm^3 (*44, 45*). (*46, 47*) This leads to an expression for n of:

$$[3] \qquad n = \frac{\left(MW_{poly}\right)\left(\rho_w\right)}{\left(MW_w\right)\left(\rho_{poly}\right)}$$

where MW_{poly} and MW_w represent the molecular weight of the polymer and solvent, respectively, and ρ_{poly} and ρ_w represent the mass density of the polymer and solvent, respectively.

The phase diagram we wished to construct consists of four, temperature-dependent volume fractions (ϕ_b', ϕ_b'', ϕ_s', and ϕ_s'') corresponding to the low-concentration arm of the binodal, the high-concentration arm of the binodal, the low-concentration arm of the spinodal, and the high-concentration arm of the spinodal, respectively.

We construct the phase diagrams using an algorithm to: 1) fit the measured saturation curve to extract a Flory χ value for each of the experimentally measured

points; 2) combine the common tangent rule (*i.e.* $\frac{d\Delta G_{mix}}{d\phi}({\phi_b}',T) = \frac{d\Delta G_{mix}}{d\phi}({\phi_b}'',T)$)

with the search for χ to produce the high concentration arm of the binodal (*i.e.* $({\phi_b}'',T)$);

and 3) use the calculated value of χ to determine the values of the spinodal for which

$\frac{d^2\Delta G_{mix}}{d\phi^2}(\phi,T) = 0$. The mathematical procedure for finding a common tangent line is

called a Maxwell construction (also known as a common tangent construction).(*48, 49*)

This algorithm is presented in pseudo-code in Figures 3 A-F. In the pseudo-code form of

the algorithm, left binodal $= \phi_b'$, right binodal $= \phi_b''$, left spinodal $= \phi_s'$, and right

spinodal $= \phi_s''$.

Given appropriate values of χ, the Flory interaction parameter, and *n*, as described

above, the free energy as a function of volume fraction exhibits a non-convex shape with

two local minima surrounding a local maximum for a given temperature. The two points

of inflection are the spinodal (ϕ_s' and ϕ_s'') and the two points where the common tangent

line meets the function are the binodal (ϕ_b' and ϕ_b''). Since the position of the left binodal

(ϕ_b') decreases monotonically as χ increases, the search can be performed using standard

root-finding algorithms. The search was implemented as a binary search described in

Figure 3.4 F) inside a subroutine that performs the Maxwell construction to obtain the

binodal. The resulting solution for χ gives the free energy function a concrete form from

which the spinodal (ϕ_s' and ϕ_s'') was derived analytically. The high-concentration arm of

the binodal (ϕ_b'') was computed using the Maxwell construction.

```
// Global definitions

n = lattice sites occupied by a single polymer; constant

// The Flory-Huggins free energy of mixing, in units of k_B T.
// χ is the Flory interaction parameter.
ΔG_mix(χ, φ) = φ(1 − φ)χ + (1/n)φ ln φ + (1 − φ) ln(1 − φ)

// The derivative of ΔG_mix with respect to φ:
dΔG_mix/dφ (χ, φ) = 1/n − 1 + χ(1 − 2φ) + (1/n) ln φ + ln(1 − φ)

// The spinodal, which is obtained analytically by taking the second
// derivative of ΔG_mix with respect to φ,
// setting the result to zero, and solving for φ_spinodal:
spinodal(χ) = (1 − n + 2nχ ± √((n − 1 − 2nχ)² − 8nχ)) / 4nχ
```

Figure 3.4 A) The first step for generating the phase diagram is defining the parameters

used in the search for χ.

```
// Solves for χ given the measured saturation concentration as a volume
// fraction

function solveForChi(saturationConcentration) {
    // If χ is less than or equal to this bound, the Flory-Huggins
    // free energy of mixing is convex and the binodal does not exist:
    loBound = √(1/n + (n+1)/(2n))
    // Substitute some safe upper bound here. If the found value of χ is less
    // than the bound, then the bound was safe:
    hiBound = 6
    // The position of the left binodal decreases monotonically as a function
    // of χ:
    return binarySearch((left binodal of maxwellConstruction(χ)), χ,
                        saturationConcentration, loBound, hiBound)
}
```

Figure 3.4 B) The second step for generating the phase diagram is to define bounds for

the values of χ in the search process. Because Equation [1] requires concentrations in

volume fraction, we calculated ϕ_b', the saturation concentration on a volume fraction

scale, by converting the measured supernatant peptide concentrations (C_{sat} on a molar

scale). This conversion was performed assuming a density for the polyglutamine

monomer (ρ_{poly}) of 1.43 g/cm^3 (*44, 45*). The conversion is: $\phi_b' = (MW_{poly})(C_{sat})/$

ρ_{poly}, where MW_{poly} is the molecular weight of the polymer. A binary search was used to

find the value of χ that yields a left binodal equal to the measured saturation

concentration (ϕ_b').

```
// Performs a Maxwell construction given a value for χ

function maxwellConstruction(χ) {
  // The slope will attain its minimum and maximum values at the spinodal:
  loBound = dΔGmix/dφ(χ, right spinodal(χ))
  hiBound = dΔGmix/dφ(χ, left spinodal(χ))
  slope = binarySearch(offset(χ, m), m, 0, loBound, hiBound)
  return findPhiWithSlope(χ, slope)
}
```

Figure 3.4 C) The Maxwell construction was implemented as a two-level nested

numerical search. The outer search finds a value of the slope (*m*) such that $\Delta G_{mix}(\phi_b'') -$

$\Delta G_{mix}(\phi_b') = m(\phi_b'' - \phi_b')$, where ϕ_b' and ϕ_b'' are, respectively, the minimum and

maximum volume fractions at which the derivative of the free energy of mixing with

respect to volume fraction equals *m*. These conditions define a common tangent line that

meets the free energy function at ϕ_b' and ϕ_b''. The possible values of *m* are bounded by

the values of the derivatives at the spinodal points. At each iteration of the outer search,

ϕ_b' and ϕ_b'' must themselves be computed using an inner numerical search over the

intervals $(0, \phi_s')$ and $(0, \phi_s'')$, respectively. The derivative, which is computed

analytically, is monotonically increasing over both of these intervals.

```
// Returns the amount by which a tangent line with slope m "misses" the
// other point with slope m. The common tangent line will nail both points,
// resulting in offset == 0

function offset($\chi$, m) {
   $\{\phi_L, \phi_R\}$ = findPhiWithSlope($\chi$,m)
   return m$(\phi_R - \phi_L) - (\Delta G_{mix}(\phi_R) - \Delta G_{mix}(\phi_L))$
}
```

Figure 3.4 D) The amount by which the common tangent line passing through the measured low-concentration arm of the binodal "misses" the predicted value of the high concentration arm of the binodal was minimized in the binary search for χ.

```
// Finds $\phi_L$ and $\phi_R$, respectively the minimum and maximum values of $\phi$ where
// $\frac{d\Delta G_{mix}}{d\phi}(\chi,\phi)$ == slope

findPhiWithSlope($\chi$, slope) {
   $\phi_L$ = binarySearch($\frac{d\Delta G_{mix}}{d\phi}(\chi,\phi)$, $\phi$, slope, 0, left spinodal($\chi$))
   $\phi_R$ = binarySearch($\frac{d\Delta G_{mix}}{d\phi}(\chi,\phi)$, $\phi$, slope, right spinodal($\chi$), 1)
   return $\{\phi_L, \phi_R\}$
}
```

Figure 3.4 E) Given a value for χ and a slope, a binary search was performed to find two concentrations on the free energy of mixing curve defined by χ corresponding to this slope.

```
// Performs a binary search to solve for the variable x such that
// func(x) == target. Assumes that x lies in the open interval
// (loBound, hiBound) and that func is monotonically increasing over
// that interval

function binarySearch(func, x, target, loBound, hiBound) {
  lo = loBound, hi = hiBound
  // In practice, it's safe to just run this loop for a constant number
  // of iterations instead of checking for convergence:
  while (hi - lo > tolerance) {
    mid = (lo + hi) / 2
    // use > instead of < here for monotonically decreasing functions
    if func(mid) < target
      lo = mid
    else
      hi = mid
    end if
  }
  return (lo + hi) / 2
}
```

Figure 3.4 F) An iterative binary search used for monotonically increasing functions.

For prediction of the entire phase diagram at multiple temperatures, it was necessary to determine χ as a function of temperature (T). This parameter is usually temperature dependent and is commonly assumed to take the form $\chi(T) = a + b/T$ *(28)*, where a and b are empirically determined by curve fitting. These parameters are thought to reflect contributions of entropy (*a*) and enthalpy (*b*) to χ.

For a given saturation curve, we used the procedure described in Figures 3 A-F) to calculate the value of χ that reproduced the measured points on the saturation curve. We then numerically fit these calculated values of χ to the form $\chi(T) = a + b/T$ to extract the parameters *a* and *b*. With knowledge of these parameters, we could calculate a value for χ at a defined temperature. We could then use this χ in equation [1] to calculate the

122

free energy of mixing for all temperatures and volume fractions. The spinodal (ϕ_s', ϕ_s'') was calculated analytically using the condition $\frac{d^2 \Delta G_{mix}}{d\phi^2}(\phi_s', T) = \frac{d^2 \Delta G_{mix}}{d\phi^2}(\phi_s'', T) = 0$. The binodal (ϕ_b', ϕ_b'') was solved using the aforementioned Maxwell construction with a known χ.

3.3.4 Atomic Force Microscopy of Polyglutamine Aggregates

To investigate the morphological development of polyglutamine aggregates, we performed AFM imaging as a function of time for constructs of $K_2Q_{30}K_2$ incubated at 40 μM at 25° C in 50 mM Phosphate Buffer at pH 7. Previously performed aggregation experiments using fluorescence anisotropy (unpublished) showed a plateau in the anisotropy value at 120 hours.

To perform the AFM imaging, 10 μL of sample was added to a freshly-cleaved mica disc attached to a glass slide. Immediately after adding the sample to the mica, 50 μL of ultrapure water was added. This dilution is important for removing the buffering salts and it also helps slow down adhesion of the protein to the mica. The diluted sample is allowed to sit on the mica for 10 seconds, at which point the liquid is removed by tilting the mica slide and pipetting the liquid away from one corner of the slide. Next, a wash step is performed by adding 100 μL of ultrapure water to the mica and then removing it as above. This wash step is repeated once more and the slide is then placed under a gentle nitrogen stream for 30 minutes to evaporate any remaining liquid. After it appears that the sample is dry, it is placed in a ventilated slide box inside a dessicant-containing oven set to 50° C. This is necessary to remove any residual water from the

sample. The sample remains in the oven for at least two hours, and is removed immediately prior to performing the experiment. The imaging was performed in tapping mode in air on an Asylum Research MFP3D Bio, and the images were collected at a 1 Hz scan rate.

3.4 Results

3.4.1 Saturation Curves for $K_2Q_{30}K_2$, $K_2Q_{40}K_2$, $Q_{30}K_2$, and $Q_{40}K_2$

The saturation curves ($C_{sat}(T)$) for the peptide constructs $KKQ_{30}KK$ (Blue Circles), $KKQ_{40}KK$ (Red Circles), $Q_{30}KK$ (Blue Triangles), and $Q_{40}KK$ (Red Triangles) are shown in Figure 3.5 A-D. Figures 3.5A and 3.5B suggest that $C_{sat}(T)$ is sensitive to the length of the polyglutamine stretch for an equivalent net charge. In particular, longer polyglutamine lengths lead to lower saturation concentrations for a given temperature. This is consistent with data from Chen et al. (*50*) although they only measured the saturation concentration at a single temperature (37˚ C).

The data support the presence of an upper critical solution temperature (UCST) which lies above the boiling point of water. Additionally, the data suggest that aqueous polyglutamine solutions may have a lower critical solution temperature (LCST), which lies below the freezing point of water. If true, this would suggest that aqueous polyglutamine solutions have a phase diagram that is best characterized as a closed-loop. (*51*) These types of phase diagrams occur when the binodal corresponding to the UCST

intersects the binodal corresponding to the LCST with the LCST being lower than the UCST. That an aqueous polyglutamine solution could have a closed-loop phase diagram is not surprising. It has been shown that aqueous solutions of amide-rich polymers which can make intrachain hydrogen bonds are capable of having a closed-loop phase diagram. (*15, 52*)
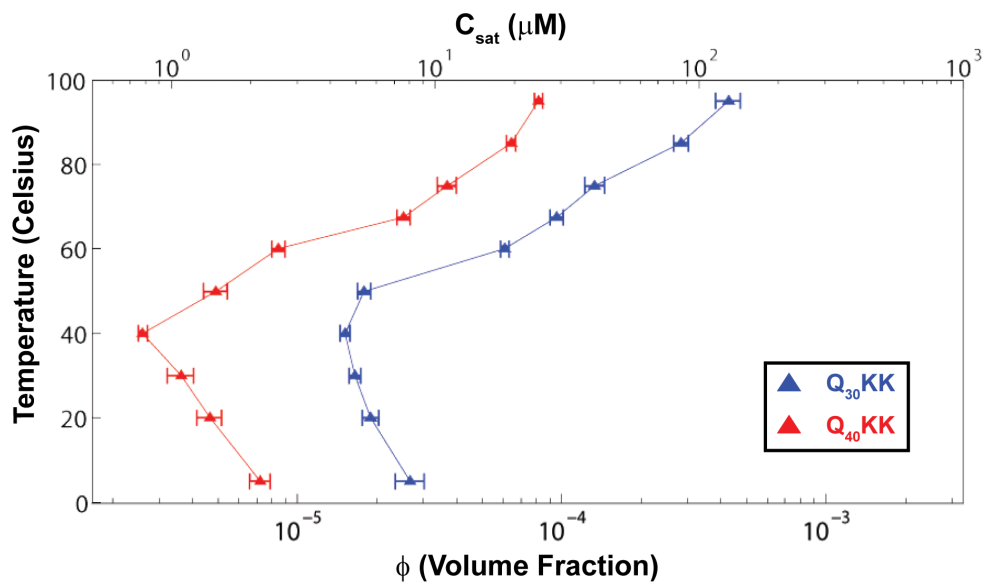
Figure 3.5 A) The saturation curves for Q$_{30}$KK (blue triangles) and Q$_{40}$KK (red triangles) plotted in a ϕ-T plane. The data suggest that the saturation concentrations are lower for longer stretches of polyglutamine. The lines are provided as a guide to the eye.
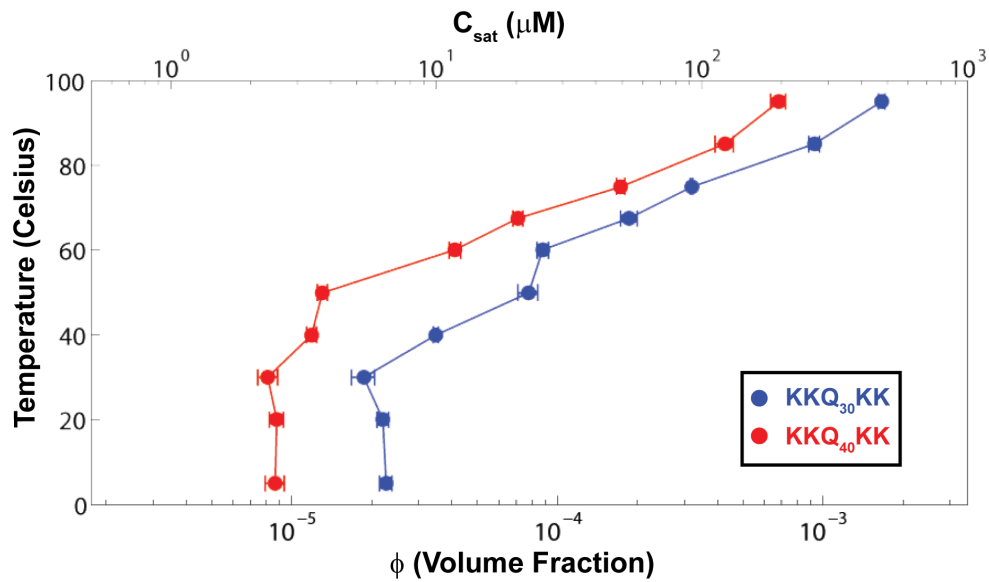
Figure 3.5 B) The saturation curves for $KKQ_{30}KK$ (blue circles) and $KKQ_{40}KK$ (red circles) plotted in a ϕ-T plane. The data suggest that the saturation concentrations are lower for longer stretches of polyglutamine. The lines are provided as a guide to the eye.

Figures 3.5C and 3.5D suggest that the net charge can also alter the saturation concentration for a given polyglutamine length. Specifically, increasing the net charge of a polyglutamine construct leads to higher saturation concentrations for a given temperature. This is consistent with observations made by Krull and Wall. (*53*)
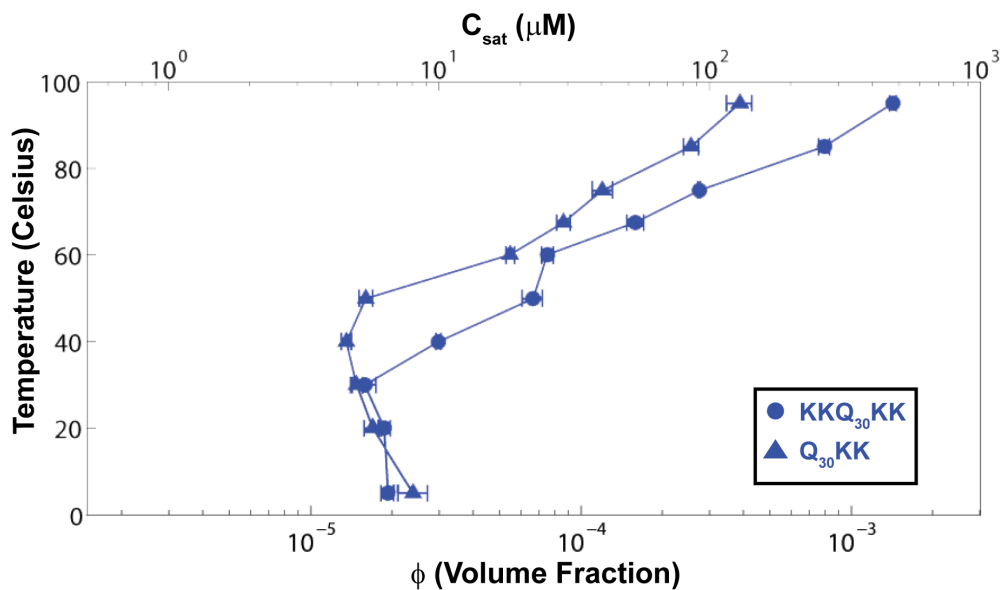
Figure 3.5 C) The saturation curves for $KKQ_{30}KK$ (blue circles) and $Q_{30}KK$ (blue triangles) plotted in a ϕ-T plane. The data suggest that the saturation concentration increases as net charge increases for a given polyglutamine length. The lines are provided as a guide to the eye.
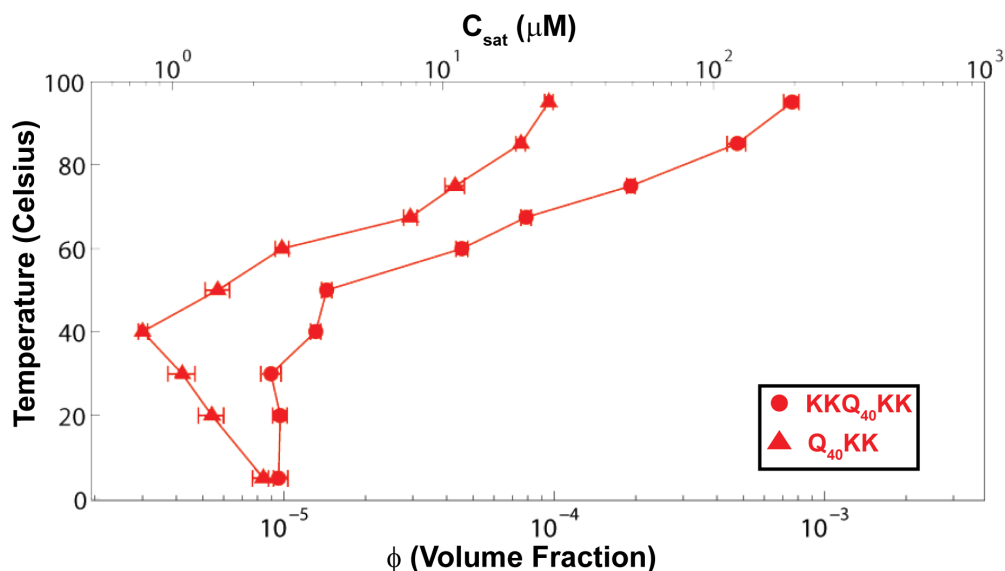
Figure 3.5 D) The saturation curves for KKQ$_{40}$KK (red circles) and Q$_{40}$KK (red triangles) plotted in a ϕ-T plane. The data suggest that the saturation concentration increases as net charge increases for a given polyglutamine length. The lines are provided as a guide to the eye.

3.4.2 Predicted Phase Diagrams

Applying the procedure described in section 3.3.3.2 to the saturation curve data from section 3.4.2, we obtained values for *a* and *b* for χ(T) shown in Table 3.1 for the different polyglutamine constructs. It should be noted that the χ values of the saturation curve that appear to be part of the binodal corresponding to an LCST were not used in the fit. Classical Flory-Huggins theory cannot be used to fit both the binodal corresponding to the UCST and the binodal corresponding to the LCST simultaneously. (*28*) It appears from the saturation curves shown in Figures 3.5 A) and 3.5B) that there is a significant

128

change in the saturation curves 40° C. Therefore, any χ corresponding to a temperature

below 40° C was not used in the fit. There appears to be a similarity in the parameters

for the constructs containing the same number of lysines.

	$KKQ_{30}KK$	$KKQ_{40}KK$	$Q_{30}KK$	$Q_{40}KK$
a	0.28	0.28	0.38	0.40
b (K)	130	134	111	99

Table 3.1 The parameters (*a* and *b*) extracted from fitting the equation $\chi(T) = a + b/T$ to

the χ values obtained for the saturation curves of $KKQ_{30}KK$, $KKQ_{40}KK$, $Q_{30}KK$, and

$Q_{40}KK$.

Using the parameters from Table 3.1 to create a temperature dependent form of

χ, we generated the phase diagrams shown in Figure 3.6 for $KKQ_{30}KK$ (Blue),

$KKQ_{40}KK$ (Red), $Q_{30}KK$ (Green), and $Q_{40}KK$ (Cyan).
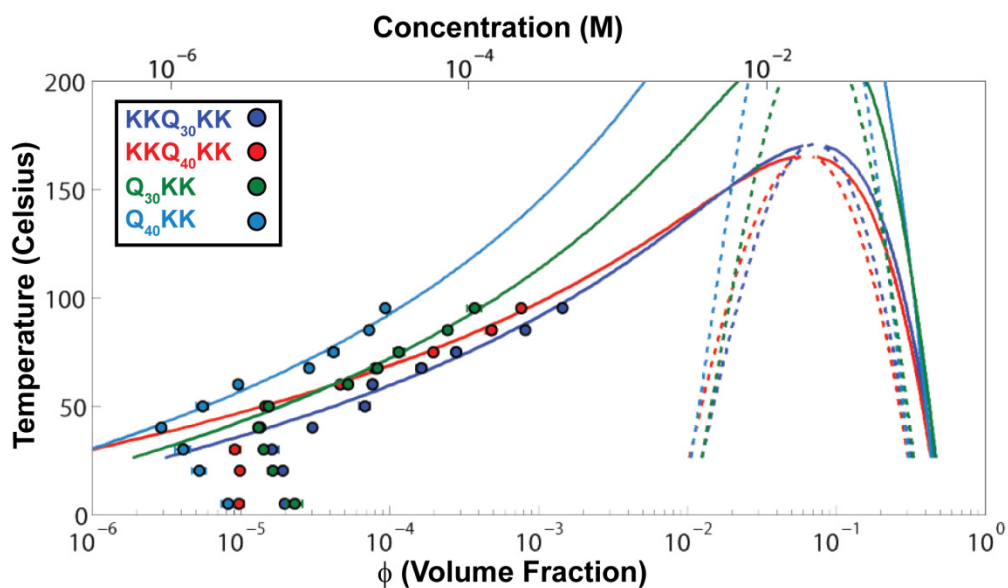
Figure 3.6 The predicted binodals (solid lines) and spinodals (dotted lines) for the peptide constructs KKQ$_{30}$KK (Blue), KKQ$_{40}$KK (Red), Q$_{30}$KK (Green), and Q$_{40}$KK (Cyan) are shown. The measured saturation concentrations that were used to predict the phase diagrams are also plotted (Circles).

3.5 Discussion

3.5.1 Implications of the inferred phase diagram

Prior to discussing the inferences we draw from the calculated phase diagrams shown in Figure 3.6, it is important to summarize the assumptions that go into the classical Flory-Huggins formalism that we have used in this work to construct a zero order phase diagram that quantifies the thermodynamic driving force for phase separation defined here by the location of the precipitation boundary.

3.5.1.1 Critique of the mean-field, random mixture Flory-Huggins model

The assumptions are as follows:

1. Each molecular entity (whether they be polymer or solvent units) has the same number of nearest neighbors z, which is encoded by the lattice model nature of the theory.

2. The interaction energy between the constituents (polymer and solvent) of the system is proportional to the product of their concentrations *viz.*, $\phi(1-\phi)$. This implies a random mixture model whereby the energy is governed by average parameters such as the bulk concentrations rather than the details of spatial arrangements of molecules with respect to each other. As a result, the effects of interactions are captured in an averaged sense.

3. Only binary interactions between nearest neighbors on the lattice are considered and this is a reasonable assumption in scenarios where the dominant intermolecular interactions are short-range ($< 10\text{-}15$ Å). Long-range interactions can have a confounding effect when using the Flory-Huggins theory. (*43*)

4. Mixing of the molecular entities does not cause a change in volume and therefore entities from both sets of molecules (polymer and solvent) can fit on the same lattice. This assumption is highly questionable and is remedied by introducing a temperature independent additive constant in the expression for χ, although this in and of itself is insufficient for addressing the weaknesses inherent to the assumption of invariant volumes. As noted earlier, an empirical parameterization

leads to $\chi(T) = a + \dfrac{b}{T}$ as a functional form that tries to separate entropic effects

due to mixing of the polymer and solvent degrees of freedom (*a*) and enthalpic

effects due intermolecular interactions (*b*).

5. Finally, Muthukumar has shown (*43*) the need for generalizing the Flory-Huggins

    theory to include the effects of conformational equilibria (*i.e.* the distinction

    between globules and coils), a generalization that we have not considered here

    because it requires that we complement the solubility measurements with separate

    measurements of the concentration-dependent correlation length ξ (requiring

    concentration-dependent measurements of scattered light) and the swelling ratio α

    (which measures the degree of swelling or compaction vis-à-vis the Flory random

    coil state). From the vantage point of precipitation, Muthukumar's formalism

    helps alleviate the difficulties that the Flory-Huggins theory faces in reproducing

    the scaling of the critical volume fraction ϕ_c with molecular weight / chain length

    parameterized as *n*.

The assumptions listed above highlight the caveats that we need to bear in mind

when interpreting the results shown in Figure 3.6. Although the assumptions are

significant and simplifying in nature, they do not preclude a comparative analysis of

results obtained for different polyglutamine lengths for a given construct and between

different constructs for peptides with similar polyglutamine lengths.

3.5.1.2 Features of the calculated phase diagrams and their implications:

The observations from the phase diagram can be summarized as follows:

1. **Saturation curves**: The measured C_{sat} values are in the micromolar range for the four peptides studied, falling below 1 μM for $Q_{40}KK$ at 40° C. For a given peptide construct and polyglutamine length, C_{sat} increases with increasing temperature and the inferred phase diagrams suggest the existence of a UCST for these systems. Fixing the temperature and peptide construct leads to lowering of C_{sat} as polyglutamine length increases. Finally, fixing the temperature and polyglutamine length shows that C_{sat} increases as the number of flanking lysine residues increase. This cannot be rationalized by a modest increase in n, but instead it points to the contributions from electrostatic repulsions that stabilize the miscible forms as is the case in colloidal suspensions.

2. **Implication of saturation curves for in vivo aggregation / phase separation**: The micromolar values of C_{sat} are confounding because it suggests that the concentrations of polyglutamine molecules *in vivo* need to be greater than C_{sat} if there is to be a driving force for aggregation that leads to phase separation. Most estimates of *in vivo* levels of polyglutamine molecules (as well as concentrations of molecules such as the amyloid beta peptide (*54-57*)) are in the nanomolar or even picomolar range. And yet, insoluble neuronal intranuclear inclusions are the defining hallmarks of polyglutamine expansion diseases.(*58-61*) This suggests the presence of one or more ways of achieving effective concentrations C_{eff} within the *in vivo* milieu that are higher than C_{sat}. From a physico-chemical perspective,

these effects might include (a) covalent modifications of the aggregation-prone molecules that lower C_{sat} (something that has been invoked for Aβ(*62, 63*)) but not for polyglutamine molecules, (b) adsorption which lowers C_{sat} through a change in dimensionality, which is achievable through interactions with membranes or other macromolecules, (c) confinement, which should cause $C_{eff} >$ C_{sat} by confining small numbers of molecules into small volumes, such as vesicles, and (d) macromolecular crowding / preferential interactions / exclusions in multicomponent systems such as cellular milieu. Of these effects, those due to confinement have been scrutinized in this work, albeit for Aβ, and the results will be discussed in Chapter 5.

3. **Metastability**: Figure 3.6 can be used to quantify the gap between the binodal and spinodal for each of the peptide systems. The results are summarized in Figure 3.7. Between the binodal and spinodal, the miscible phase is metastable because $\dfrac{\partial^2 \Delta \bar{G}_{mix}}{\partial \phi^2} > 0$ in this region even though there is a thermodynamic driving force for phase separation and therefore we refer to the gap between the binodal and spinodal as the width of the metastable region or Δ_M defined as:

$$[4] \quad \Delta_M = \log_{10}(\phi_{spinodal}) - \log_{10}(\phi_{binodal})$$

Figure 3.7 makes several points. First, Δ_M can span up to four orders of magnitude for the lowest temperatures, with this value being greater than 3 at 37° C. Second, Δ_M narrows as temperature increases. This suggests that the

metastability is more pronounced at lower temperatures, particularly at temperatures that are of physiological relevance. The kinetics of phase separation in all likelihood will follow pathways that pass through metastable states because these will involve lower free energy barriers than a direct transition to the equilibrium stable state via a high free energy barrier corresponding to a singular bottleneck. The larger the gap Δ_M, the greater the likelihood of accessing these metastable states. Detecting these metastable states requires that they are long lived. It also requires that each metastable state (or mesophase in the physics literature) be distinguishable in terms of a relevant order parameter such as density or composition for using Flory-Huggins theory. Deficiencies in detection technologies do not preclude the existence of mesophases in the metastable region (*64*). In fact, our calculated phase diagram suggests that Δ_M becomes more pronounced as polyglutamine length increases for a given construct. The metastable states or mesophases could well be the heterogeneous distributions of oligomers that have been characterized by Muchowski and coworkers (*65-67*) as well as several other groups for different aggregation-prone systems (*68-73*). In the next chapter we will visit the tenets of classical nucleation theory and analyze published results for aggregation kinetics to demonstrate that Δ_M values inferred here are consistent with a mechanism of phase separation that proceeds through oligomers / clusters that coalesce first, then sprout through a barrier-limited (nucleated) conformational conversion within the oligomers followed by a slow ripening and condensation to form long fibrillar structures. This mode of phase

135

separation or "assembly" appears to be widely accepted in the protein aggregation literature although the prominent biophysical model for polyglutamine systems is the homogeneous nucleation model of Wetzel and coworkers. (*18, 19, 74*) In Section 3.5.2 we will show that the inferred Δ_M values are consistent with a mechanistic description laid out above.
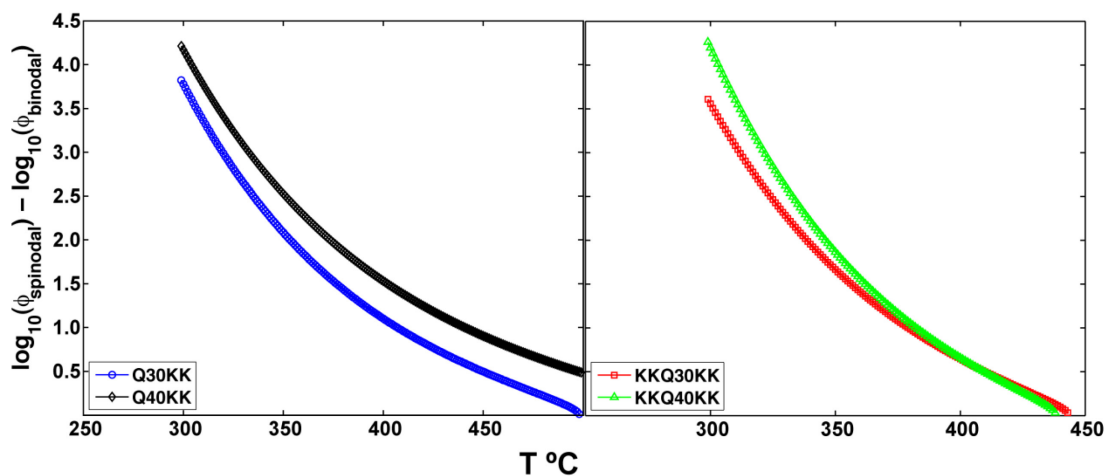
Figure 3.7 Widths of the metastability regions Δ_M plotted against increasing temperature for each of the peptides that were part of the measurement sets.

3.5.2 Testing metastability inferred from the calculated phase diagrams

The incubation times for solubility measurements were greater than two weeks. Previous fluorescence anisotropy measurements on the time-course of aggregation suggested that the process was complete for KKQ$_{30}$KK in 120 hours for 40 μM solutions incubated at 25° C under identical solution conditions used in the determination of the

136

saturation curves in Section 3.4.1. Under these conditions, the width of the metastability gap (Δ_M) is near its maximum value, suggesting a high likelihood that aggregation will proceed through heterogeneous species.

We used AFM measurements as described in section 3.3.4 to make assessments of the morphology of species populated during aggregation under the aforementioned conditions. Small aliquots were taken at 120 hours and 24 hours later. The results are shown in Figures 3.8 A and B.
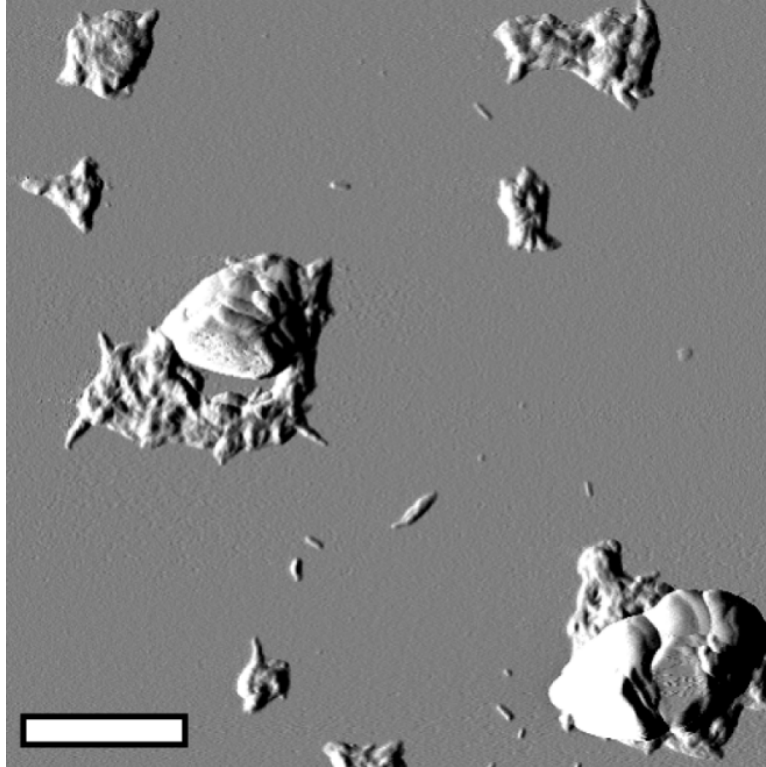
Figure 3.8 A) AFM image of a 40 μM solution of KKQ$_{30}$KK incubated at 25° C for 120 hours in 50 mM phosphate buffer at pH 7. Previously performed fluorescence anisotropy experiments suggested that aggregation should be complete in 120 hours, but the globular, amorphous morphologies seen here are inconsistent with the expected final form, which is a linear fibril. (*17, 75*) We see large amorphous aggregates which are on the order of 1 μm wide and 100 nm tall. The scale bar is 1 μm.
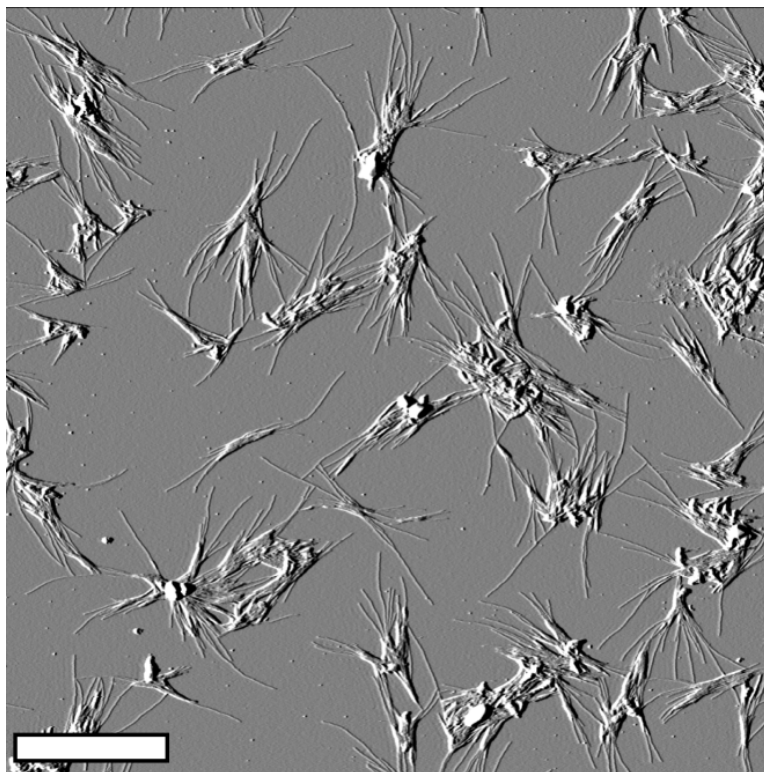
Figure 3.8 B) AFM image of a 40 µM solution of $KKQ_{30}KK$ incubated at 25° C for 144 hours in 50 mM phosphate buffer at pH 7. This image was taken 24 hours after the image in Figure 3.A. One can see that the amorphous globules in Figure 3.8 A have converted by "sprouting" fibrils from their centers. This is consistent with nucleation of the fibril phase occurring inside the liquid-like aggregates seen in Figure 3.8 A. The scale bar is 1 µm.

At 120 hours, we see large amorphous aggregates (Figure 3.8A) which are on the order of 1 µm wide and 100 nm tall. The same sample imaged 24 hours later (Figure 3.8B) shows that the aggregates are no longer amorphous. Instead, it appears that the

139

amorphous aggregates have converted into species with thin fibrils (<10 nm in width and height) emanating from what appear to be amorphous centers.

KKQ$_{30}$KK has the lowest aggregation propensity based on the measured saturation curves (Figures 3.5 A and B), of any of the polyglutamine constructs measured. At the other end of the spectrum, Q$_{40}$KK has the highest aggregation propensity. Q$_{40}$KK also has a higher metastability gap at 25° C than KKQ$_{30}$KK (Figure 3.7). These points suggest that aggregation of Q$_{40}$KK should proceed through more heterogeneous intermediates, and it should occur faster. We used AFM measurements as described in section 3.3.4 to make assessments of the morphology of species populated during aggregation of a sample of Q$_{40}$KK incubated at 25° C in 50 mM Phosphate buffer (pH 5). Previously performed aggregation experiments using fluorescence anisotropy (unpublished) showed that the anisotropy value reached a plateau within 15 hours. Small aliquots were removed as soon as the solution was resuspended in aqueous milieu, at 20 minutes after re-suspension, and at 45 minutes. The AFM images obtained from these aliquots are shown in Figures 3.9 A-C.
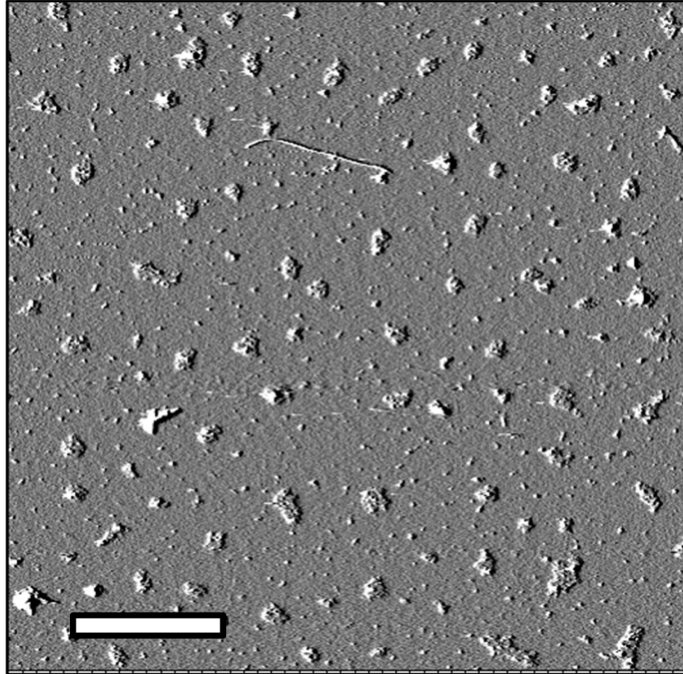
Figure 3.9 A) AFM image of a solution of 40 μM $Q_{40}KK$ was resuspended in 50 mM phosphate buffer at pH 5 and imaged immediately. This image represents the earliest time point which can be observed in our experiments. The image show a heterogeneous distribution of amorphous aggregates form almost instaneously upon re-suspension in an aqueous milieu. The scale bar is 1 μm.
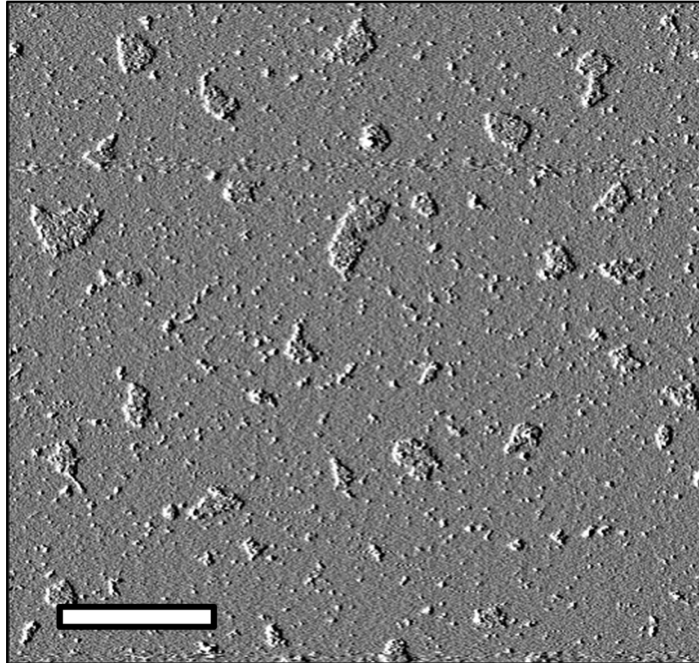
Figure 3.9 B) AFM image of a solution of 40 μM $Q_{40}KK$ incubated at 25° C for 20 minutes in 50 mM phosphate buffer at pH 5. Compared to Figure 3.9 A), one can see that amorphous aggregates have begun to coalesce into larger structures. The scale bar is 1 μm.
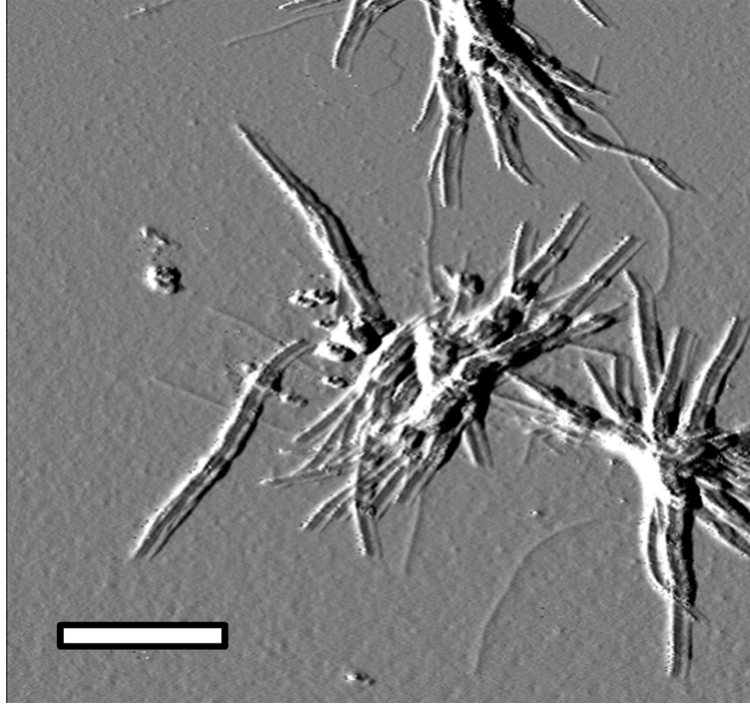
Figure 3.9 C) AFM image of a solution of 40 μM $Q_{40}KK$ incubated at 25° C for 45 minutes in 50 mM phosphate buffer at pH 5. Compared to Figures 3.9 A) and B), the heterogeneous puncta appear to have essentially disappeared. In their place, fibril-like aggregates appear to be emanating from amorphous centers. The fibrils appear to have a shadowy appearance. This is due to the presence of low amounts of water. This water is difficult to remove due to the high relative humidity in the summer, which is when these experiments were performed. The scale bar is 1 μm.

The aggregation of $Q_{40}KK$ as assessed by AFM proceeds very rapidly. A heterogeneous distribution of small aggregates (~10-100 nm in width and ~10-50 nm in height) were present as quickly as the sample could be plated (< 1 min) after re-

suspension in buffer (Figure 3.9 A). Within 20 minutes, the small aggregates had begun

to coalesce into larger aggregates (~100-500 nm in width and ~10-50 nm in height), but

some small aggregates still remained (Figure 3.9 B). Within 45 minutes, fibrillar

structures (> 1 mm in length and > 10 nm in height) appeared (Figure 3.9 C). The fibrils

appeared at the expense of smaller aggregates as these are not observed in any reasonable

amount compared to Figures 3.9 A and B. This is consistent with the prediction that

formation of the final phase proceeds through heterogeneous species.

A comparison of the AFM images for the aggregation of $KKQ_{30}KK$ (Figures 3.8 A

and B) and $Q_{40}KK$ (Figures 3.9 A-C) shows that the aggregation of both constructs

proceeds through heterogeneous intermediates before reaching the final fibrillar state.

The aggregation of $Q_{40}KK$ is more rapid than $KKQ_{30}KK$, consistent with previously

performed fluorescence anisotropy experiments. The insoluble phase for $Q_{40}KK$ is

reached within 45 minutes as compared to more than 120 hours for $KKQ_{30}KK$ given the

same starting peptide concentration (40 μM). Both of these observations are consistent

with the phase diagram shown in Figure 3.6.

3.6 Summary

We have used techniques and theories from classical polymer physics to construct a phase diagram for aqueous solutions of polyglutamine constructs $KKQ_{30}KK$, $KKQ_{40}KK$, $Q_{30}KK$, and $Q_{40}KK$. These phase diagrams provide thermodynamic constraints on the aggregation mechanism. In particular, they suggest that the formation of the fibrillar phase will occur through nucleation and growth under normal experimental conditions. Importantly, however, due to the large gap between the low concentration arm of the binodal and the calculated spinodal, the aggregation is likely to proceed through the formation of heterogeneous species in which nucleation of the fibrillar phase is likely to occur. Atomic Force Microscopy images are consistent with this hypothesis.

3.7 References

1.	Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease, *Annu Rev Biochem 75*, 333-366.
2.	Meirovitch, H., and Scheraga, H. A. (1980) Empirical Studies of Hydrophobicity. 2. Distribution of the Hydrophobic, Hydrophilic, Neutral, and Ambivalent Amino Acids in the Interior and Exterior Layers of Native Proteins, *Macromolecules 13*, 1406-1414.
3.	Tanford, C. (1978) The Hydrophobic Effect and the Organization of Living Matter, *Science 200*, 1012-1018.
4.	Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly, *Nature 437*, 640-647.

5.   Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) Rationalization of the effects of mutations on peptide andprotein aggregation rates, *Nature 424*, 805-808.

6.   Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G., and Dobson, C. M. (2002) Kinetic partitioning of protein folding and aggregation, *Nat Struct Biol 9*, 137-143.

7.   Kiefhaber, T., Rudolph, R., Kohler, H.-H., and Buchner, J. (1991) Protein Aggregation in vitro and in vivo: A Quantitative Model of the Kinetic Competition between Folding and Aggregation, *Nat Biotech 9*, 825-829.

8.   Pashley, R., McGuiggan, P., Ninham, B., and Evans, D. (1985) Attractive forces between uncharged hydrophobic surfaces: direct measurements in aqueous solution, *Science 229*, 1088-1089.

9.   Rank, J. A., and Baker, D. (1997) A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding, *Protein Science 6*, 347-354.

10.   Wiggins, P. M. (1997) Hydrophobic hydration, hydrophobic forces and protein folding, *Physica A: Statistical and Theoretical Physics 238*, 113-128.

11.   Lesk, A. M., and Chothia, C. (1980) Solvent accessibility, protein surfaces, and protein folding, *Biophysical journal 32*, 35-47.

12.   Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R., and Pappu, R. V. (2006) Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions, *Proc Natl Acad Sci U S A 103*, 16764-16769.

13.   McNaught, A. D., Wilkinson, A., and International Union of Pure and Applied Chemistry. (1997) *Compendium of chemical terminology : IUPAC recommendations*, 2nd ed., Blackwell Science, Oxford England ; Malden, MA, USA.

14.   Rose, G. D., and Wolfenden, R. (1993) Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding, *Annual Review of Biophysics and Biomolecular Structure 22*, 381-415.

15.   Bokias, G., Staikos, G., and Iliopoulos, I. (2000) Solution properties and phase behaviour of copolymers of acrylic acid with N-isopropylacrylamide: the importance of the intrachain hydrogen bonding, *Polymer 41*, 7399-7405.

16.   Holthauzen, L. M., Rosgen, J., and Bolen, D. W. (2010) Hydrogen bonding progressively strengthens upon transfer of the protein urea-denatured state to water and protecting osmolytes, *Biochemistry 49*, 1310-1318.

17.   Chen, S., Berthelier, V., Hamilton, J. B., O'Nuallain, B., and Wetzel, R. (2002) Amyloid-like features of polyglutamine aggregates and their assembly kinetics, *Biochemistry 41*, 7391-7399.

18.   Chen, S., Ferrone, F. A., and Wetzel, R. (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation, *Proc Natl Acad Sci U S A 99*, 11884-11889.

146

19. Kar, K., Jayaraman, M., Sahoo, B., Kodali, R., and Wetzel, R. (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent, *Nat Struct Mol Biol 18*, 328-336.

20. Dobry, A., and Boyer-Kawenoki, F. (1947) Phase separation in polymer solution, *Journal of Polymer Science 2*, 90-100.

21. De Young, L. R., Fink, A. L., and Dill, K. A. (1993) Aggregation of globular proteins, *Accounts of Chemical Research 26*, 614-620.

22. Ciferri, A. (2005) *Supramolecular Polymers*, CRC Press.

23. Flory, P. J. (1953) *Principles of Polymer Chemistry*, Cornell University Press, Ithaca and London.

24. Grosberg, A. Y., and Khokhlov, A. R. (1997) *Statistical Physics of Macromolecules (AIP Series in Polymers and Complex Materials)*, 1 ed., AIP Press, New York.

25. Grosberg, A. Y., and Kuznetsov, D. V. (1992) Phase-Separation Of Polymer-Solutions And Interactions Of Globules, *Journal De Physique Ii 2*, 1327-1339.

26. Pappu, R. V., Wang, X., Vitalis, A., and Crick, S. L. (2008) A polymer physics perspective on driving forces and mechanisms for protein aggregation, *Arch Biochem Biophys 469*, 132-141.

27. Raos, G., and Allegra, G. (1996) Chain collapse and phase separation in poor-solvent polymer solutions: A unified molecular description, *Journal Of Chemical Physics 104*, 1626-1645.

28. Rubinstein, M., and Colby, R. H. (2003) *Polymer Physics*, Oxford University Press, Oxford and New York.

29. Anderson, V. J., and Lekkerkerker, H. N. W. (2002) Insights into phase transition kinetics from colloid science, *Nature 416*, 811-815.

30. Cosgrove, T. (2010) *Colloid Science: Principles, Methods, and Applications*, Second ed., John Wiley & Sons Ltd.

31. Kruyt, H. R. (1959) *Colloid Science*, Vol. 1, Elsevier.

32. Verwey, E. J. W., and Overbeek, J. T. G. (1948) *Theory of the stability of lyophobic colloids*, Elsevier, Amsterdam.

33. Tavares, F. W., and Sandler, S. I. (1997) Phase equilibria for the mean-force potential of globular protein solutions, *AIChE Journal 43*, 218-231.

34. Muschol, M., and Rosenberger, F. (1997) Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization, *Journal of Chemical Physics 107*, 1953-1962.

35. Asherie, N. (2004) Protein crystallization and phase diagrams, *Methods 34*, 266-272.

36. Galkin, O., and Vekilov, P. G. (1999) Are Nucleation Kinetics of Protein Crystals Similar to Those of Liquid Droplets?, *Journal of the American Chemical Society 122*, 156-163.

37. Wetzel, R. (2005) In *Protein Folding Handbook* (Buchner, J., and Kiefhaber, T., Eds.), Wiley-VSH Verlag GmbH & Co Weinheim.

38. Chen, S., and Wetzel, R. (2001) Solubilization and disaggregation of polyglutamine peptides, *Protein Sci 10*, 887-891.

39. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid, *Anal Biochem 150*, 76-85.

40. Van Norman, K. H. (1909) The Biuret Reaction and the Cold Nitric Acid Test in the Recognition of Protein, *Biochem J 4*, 127-135.

41. Flory, P. J. (1941) Thermodynamics of High Polymer Solutions, *The Journal of Chemical Physics 9*, 660-660.

42. Huggins, M. L. (1941) Solutions of Long Chain Compounds, *The Journal of Chemical Physics 9*, 440-440.

43. Muthukumar, M. (1986) Thermodynamics of polymer solutions., *Journal of Chemical Physics 85*.

44. Fischer, H., Polikarpov, I., and Craievich, A. F. (2004) Average protein density is a molecular-weight-dependent function, *Protein Science 13*, 2825-2828.

45. Quillin, M. L., and Matthews, B. W. (2000) Accurate calculation of the density of proteins, *Acta Crystallographica Section D 56*, 791-794.

46. Fields, G. B., Alonso, D. O. V., Stigter, D., and Dill, K. A. (1992) Theory for the aggregation of proteins and copolymers, *The Journal of Physical Chemistry 96*, 3974-3981.

47. Nagarajan, R., and Wang, C.-C. (2000) Theory of Surfactant Aggregation in Water/Ethylene Glycol Mixed Solvents, *Langmuir 16*, 5242-5251.

48. Ge, H., and Qian, H. (2009) Thermodynamic limit of a nonequilibrium steady state: Maxwell-type construction for a bistable biochemical system, *Phys Rev Lett 103*, 148103.

49. Parola, A., Pini, D., and Reatto, L. (1993) First-order phase transitions, the Maxwell construction, and the momentum-space renormalization group, *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 48*, 3321-3332.

50. Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001) Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity, *J. Mol. Biol. 311*, 173-182.

51. Qian, C. B., Mumby, S. J., and Eichinger, B. E. (1991) Phase-Diagrams of Binary Polymer-Solutions and Blends, *Macromolecules 24*, 1655-1661.

52. Rebelo, L. P. N., Visak, Z. P., de Sousa, H. C., Szydlowski, J., Gomes de Azevedo, R., Ramos, A. M., Najdanovic-Visak, V., Nunes da Ponte, M., and Klein, J. (2002) Double Critical Phenomena in (Water + Polyacrylamides) Solutions, *Macromolecules 35*, 1887-1895.

53. Krull, L. H., and Wall, J. S. (1966) Synthetic Polypeptides Containing Side-Chain Amide Groups. Water-Soluble Polymers*, *Biochemistry 5*, 1521-1527.

54. Strozyk, D., Blennow, K., White, L. R., and Launer, L. J. (2003) CSF Abeta 42 levels correlate with amyloid-neuropathology in a population-based autopsy study, *Neurology 60*, 652-656.

55. Cirrito, J. R., May, P. C., O'Dell, M. A., Taylor, J. W., Parsadanian, M., Cramer, J. W., Audia, J. E., Nissen, J. S., Bales, K. R., Paul, S. M., DeMattos, R. B., and Holtzman, D. M. (2003) In vivo assessment of brain interstitial fluid with

microdialysis reveals plaque-associated changes in amyloid-beta metabolism and half-life, *J Neurosci 23*, 8844-8853.

56. Andreasen, N., Hesse, C., Davidsson, P., Minthon, L., Wallin, A., Winblad, B., Vanderstichele, H., Vanmechelen, E., and Blennow, K. (1999) Cerebrospinal fluid beta-amyloid(1-42) in Alzheimer disease: differences between early- and late-onset Alzheimer disease and stability during the course of disease, *Archives of neurology 56*, 673-680.

57. Galasko, D., Chang, L., Motter, R., Clark, C. M., Kaye, J., Knopman, D., Thomas, R., Kholodenko, D., Schenk, D., Lieberburg, I., Miller, B., Green, R., Basherad, R., Kertiles, L., Boss, M. A., and Seubert, P. (1998) High cerebrospinal fluid tau and low amyloid beta42 levels in the clinical diagnosis of Alzheimer disease and relation to apolipoprotein E genotype, *Archives of neurology 55*, 937-945.

58. Hayashi, Y., Kakita, A., Yamada, M., Koide, R., Igarashi, S., Takano, H., Ikeuchi, T., Wakabayashi, K., Egawa, S., Tsuji, S., and Takahashi, H. (1998) Hereditary dentatorubral-pallidoluysian atrophy: detection of widespread ubiquitinated neuronal and glial intranuclear inclusions in the brain, *Acta Neuropathologica 96*, 547-552.

59. Cummings, C. J., and Zoghbi, H. Y. (2000) Trinucleotide repeats: mechanisms and pathophysiology, *Annu Rev Genomics Hum Genet 1*, 281-328.

60. Yamada, M., Tsuji, S., and Takahashi, H. (2000) Pathology of CAG repeat diseases, *Neuropathology 20*, 319-325.

61. Yamada, M., Sato, T., Tsuji, S., and Takahashi, H. (2008) CAG repeat disorder models and human neuropathology: similarities and differences, *Acta Neuropathologica 115*, 71-86.

62. Gunn, A. P., Masters, C. L., and Cherny, R. A. (2010) Pyroglutamate-Abeta: Role in the natural history of Alzheimer's disease, *Int J Biochem Cell Biol 42*, 1915-1918.

63. Usui, K., Hulleman, J. D., Paulsson, J. F., Siegel, S. J., Powers, E. T., and Kelly, J. W. (2009) Site-specific modification of Alzheimer's peptides by cholesterol oxidation products enhances aggregation energetics and neurotoxicity, *Proceedings of the National Academy of Sciences of the United States of America 106*, 18563-18568.

64. Frieden, C. (2007) Protein aggregation processes: In search of the mechanism, *Protein Sci 16*, 2334-2344.

65. Miller, J., Rutenber, E., and Muchowski, P. J. (2009) Polyglutamine dances the conformational cha-cha-cha, *Structure 17*, 1151-1153.

66. Legleiter, J., Mitchell, E., Lotz, G. P., Sapp, E., Ng, C., DiFiglia, M., Thompson, L. M., and Muchowski, P. J. (2010) Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner in vitro and in vivo, *J Biol Chem 285*, 14777-14790.

67. Sathasivam, K., Lane, A., Legleiter, J., Warley, A., Woodman, B., Finkbeiner, S., Paganetti, P., Muchowski, P. J., Wilson, S., and Bates, G. P. (2010) Identical

oligomeric and fibrillar structures captured from the brains of R6/2 and knock-in mouse models of Huntington's disease, *Hum Mol Genet 19*, 65-78.

68. Ceru, S., Kokalj, S. J., Rabzelj, S., Skarabot, M., Gutierrez-Aguirre, I., Kopitar-Jerala, N., Anderluh, G., Turk, D., Turk, V., and Zerovnik, E. (2008) Size and morphology of toxic oligomers of amyloidogenic proteins: a case study of human stefin B, *Amyloid 15*, 147-159.

69. Glabe, C. G. (2008) Structural classification of toxic amyloid oligomers, *J Biol Chem 283*, 29639-29643.

70. Haass, C., and Selkoe, D. J. (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide, *Nat Rev Mol Cell Biol 8*, 101-112.

71. Huang, T. H., Yang, D. S., Plaskos, N. P., Go, S., Yip, C. M., Fraser, P. E., and Chakrabartty, A. (2000) Structural studies of soluble oligomers of the Alzheimer beta-amyloid peptide, *J Mol Biol 297*, 73-87.

72. Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis, *Science 300*, 486-489.

73. Juarez, J., Taboada, P., and Mosquera, V. (2009) Existence of different structural intermediates on the fibrillation pathway of human serum albumin, *Biophys J 96*, 2353-2370.

74. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction, *Proc Natl Acad Sci U S A 102*, 15400-15405.

75. Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases, *Proc Natl Acad Sci U S A 91*, 5355-5358.

Chapter 4: Evaluating the Homogeneous Nucleation Hypothesis for Phase Separation in Polyglutamine Solutions

4.1 Preamble

Most *in vitro* aggregation experiments with polyglutamine systems are performed in the range of 1-100 μM and at either 25° C or 37° C. (*1-6*) From the phase diagrams presented in Chapter 3, it is clear that these conditions place the solution in the metastable regime where phase separation is likely to occur by a nucleation and growth mechanism. (*7*) However, the width of the metastable region can be as large as 4 orders of magnitude, suggesting that there is a high likelihood that the nucleation will be heterogeneous. This is important because the most prominent biophysical model for aggregation of polyglutamine systems is the homogeneous nucleation model of Wetzel and coworkers. (*1, 4, 5*)

This chapter focuses on classical nucleation theory (CNT), its connection between kinetics and thermodynamics, and its application to understanding the phase separation of solutions of polyglutamine. Interpretations from kinetics experiments that lead to the proposal that polyglutamine aggregation is a homogeneous nucleation process with a nucleus size of one (*5*) will be reassessed with CNT.

4.2 Introduction

4.2.1 Homogeneous nucleation and polyglutamine aggregation

It is commonly assumed that polyglutamine aggregation proceeds through homogeneous nucleation as was discussed in Chapter 1. (*1, 4, 5*) This means that there is a single free energy barrier which must be crossed before phase separation proceeds downhill. (*8-11*) In classical nucleation theory (CNT), the magnitude of this barrier is related to the size of a critical cluster (a nucleus) of protein molecules. (*9-11*) The formation of this nucleus is unfavorable due to the free energy cost of creating this cluster inside the bulk phase, which is usually the soluble phase for protein systems.

Wetzel and coworkers applied a microstate partitioning model known as a thermodynamic nucleus model (*12*) to interpret kinetic data of polyglutamine aggregation. (*1, 4, 5*) This model was discussed in detail in *Scheme 4* presented in Chapter 1, Section 1.3.5.2. They asserted, based on interpretation of their kinetic data with this model, that polyglutamine aggregation is a homogeneous nucleation and that the nucleus size is one molecule. (*1, 4*) They interpret this to mean that there is a single, rare conformation polyglutamine can adopt that will allow the formation of the new phase. If the mechanism is homogeneous nucleation, evaluation of the same data using CNT should result in similar interpretation.

4.2.2 Overview of Classical Nucleation Theory (CNT)

The simplest theory for describing the physics of nucleation is classical nucleation theory (CNT). (*9-11, 13, 14*) Before providing an evaluation of polyglutamine aggregation with CNT, it is important to provide an overview of the physics of CNT.

Nucleation is an activated process which means that the free energy barrier for forming the nucleus contributes to the rate of nucleation (as does the diffusion of molecules between phases). We will refer to this barrier as ΔF^*. The probability of crossing this barrier increases with increasing kT, the product of Boltzmann's constant and absolute temperature. For homogeneous nucleation, CNT expresses the nucleation rate per unit volume as:

$$[1] \quad rate = \rho_n Z j \exp\left(\frac{-\Delta F^*}{kT}\right)$$

where the prefactor is a product of ρ_n, the number of possible nucleation sites per unit volume (this is also referred to as a number density or molecular concentration in the literature), Z, the Zeldovich factor (the probability that a nucleus at the top of the nucleation barrier continues on to form the new phase), and, j, the growth rate of the nucleus. For homogeneous nucleation in a protein solution, one assumes that ρ_n is equal to the number of protein molecules in solution per unit volume. For heterogeneous nucleation with a single barrier, ρ_n is equal to the number of impurities in the solution per unit volume. Unless explicitly specified, the following discussion will concern homogeneous nucleation.

If the critical nucleus is approximated as a sphere of radius, R^*, then the maximal growth rate of the nucleus is the diffusion-limited rate given by $j \sim \rho_b D R^*$, where D is the diffusion coefficient for a single molecule, and ρ_b is the number density of molecules in the bulk. The rate at which the nucleus actually crosses the nucleation barrier and grows into a new phase is given by $Z j$.

The free energy of creating a nucleus with radius R of a new phase inside the bulk phase is expressed as:

$$[2] \quad \Delta F = \frac{-4\pi}{3} R^3 \rho_n \Delta\mu + 4\pi R^2 \gamma$$

where $\Delta\mu = \mu_{old} - \mu_{new} > 0$ is the chemical potential of the bulk phase minus the new, and γ is the interfacial tension between the bulk phase and the new phase.

For nucleation in a dilute, impurity-free solution, one can treat the solution as an ideal gas where the degree of supersaturation (S) is defined as:

$$[3] \quad \frac{\Delta\mu}{kT} = \ln\left(\frac{C_0}{C_{sat}}\right) \equiv S$$

where k is Boltzmann's constant, C_0 is the bulk concentration of protein, and C_{sat} is the saturation concentration. Equation [3] completely satisfies the definition of saturation concentration (*i.e.* the chemical potential difference between the bulk phase and the nucleating phase will be zero when $C_0 = C_{sat}$). For any $C_0 < C_{sat}$, the chemical potential will be negative, meaning that the soluble phase is stable. Equation [3] provides the link

between equilibrium thermodynamics of phase separation processes, and kinetics governing the processes.

Recall from Chapter 3, that we determined the saturation curves, $C_{sat}(T)$, for 4 different polyglutamine constructs (KKQ$_{30}$KK, KKQ$_{40}$KK, Q$_{30}$KK, and Q$_{40}$KK). Based on these saturation curves, one can predict from CNT that, for the same initial concentration of protein (C_0) and solution conditions, the rate of aggregation of the polyglutamine constructs at 37° C should follow the trend: Q$_{40}$KK > KKQ$_{40}$KK ≈ Q$_{30}$KK > KKQ$_{30}$KK, if homogeneously nucleation is applicable to the system. This is just an illustration of the predictive power that can be realized by connecting the thermodynamics of phase behavior to kinetics.

For the rate of nucleation, we are only concerned about the free energy maximum (ΔF^*), which is found by setting the derivative of $\Delta F = 0$ and solving for R^*. This leads to a critical radius:

[4] $\quad R^* = \dfrac{2\gamma}{(\rho_n \Delta\mu)}$

for which the nucleation barrier is:

[5] $\quad \Delta F^* = \dfrac{16\pi}{3}\dfrac{\gamma^3}{(\rho_n \Delta\mu)^2} = \dfrac{16\pi}{3(kT)^2}\dfrac{\Omega^2\gamma^3}{S^2}$

where Ω is the volume of the protein molecule in the new phase. From Equation 5, one can see that ΔF^* will tend toward infinity as S approaches 0.

The nucleation theorem relates the nucleus size ($n*$) to the change of $\Delta F *$ as a function of μ. The derivation is shown in equation [6] where the volume of the critical nucleus, $\frac{4\pi}{3}(R*)^3$, multiplied by the number density, ρ_n, gives the number of molecules in the critical nucleus, $n*$.

$$[6] \quad \frac{d\Delta F *}{d\mu} = \frac{-32\pi}{3}\frac{\gamma^3}{\rho_n^2}\left(\frac{1}{\Delta\mu}\right)^3 = \frac{-32\pi}{3}\frac{\gamma^3}{\rho_n^2}\left(\frac{R*\rho_n}{2\gamma}\right)^3 = \frac{-4\pi}{3}(R*)^3 \rho_n = -n*$$

For homogeneous nucleation, with $\frac{\Delta\mu}{kT} = S$, one can rewrite equation [6] as:

$$[7] \quad \frac{d\ln(rate)}{dS} = -\frac{d\Delta F */kT}{dS} = n*$$

Equation [7] is useful experimentally because it allows one to estimate the nucleus size as a function of a macroscopic observable, the nucleation rate. By measuring the nucleation rate over a range of concentrations, and, hence, over a range of degrees of supersaturation, one can use equation [7] to estimate $n*$.

Equation [7] technically applies to both homogeneous and heterogeneous nucleation. However, it is important to note, especially for the remainder of this chapter, that if equation [7] is used to analyze a heterogeneous nucleation process, then $n*$ will reflect a weighted average of the various nuclei's contributions to the measured nucleation rate. Moreover, the individual nuclei might all have different activation barriers. For these reasons, applying Equation [7] to a heterogeneous nucleation process results in values for $n*$ which are not interpretable.

For analysis of experimental data, the homogeneous nucleation rate, [1], is usually rewritten as:

[8] $\quad rate = A\exp\left(\dfrac{-B}{S^2}\right) \quad$ or $\quad \ln(rate) = \ln(A) - \ln\left(\dfrac{-B}{S^2}\right)$

where A and B are empirically determined from fitting nucleation rates as a function of S. Upon comparison of equations [8], [1], and [5], one can see that A and B both have a physical interpretation. The parameter, B, is related to the thermodynamic barrier for forming the critical cluster (see Equation [9]). In particular, B is a measure of the energetic cost of creating a cavity of volume (Ω) inside a liquid with an interfacial tension (γ) between the new phase and the bulk. A, is a kinetic parameter related to the nucleus stabilization and growth (see Equation [10]).

[9] $\quad B = \dfrac{16\pi}{3}\dfrac{\Omega^2\gamma^3}{(kT)^3}$

[10] $\quad A = \rho_n Z j$

By substituting equation [8] into [7], one can see that it is also possible to estimate the size of the nucleus from B:

[11] $\quad n^* = \dfrac{2B}{S^3}$

Equation [11] shows that the nucleus size is not fixed. In fact, the nucleus size will depend strongly on the degree of supersaturation for homogeneous nucleation. Equation [11] is technically true for both homogeneous and heterogeneous nucleation, as long as there is only a single nucleation barrier. When heterogeneities are introduced, it

is likely that there will be a distribution of nucleation barriers, and, hence, a distribution of nuclei. Importantly, for a homogenous nucleation process, the nucleus size determined from equation [11] and equation [7] will be the same.

Let us examine the case of heterogeneous nucleation on impurities. We will assume that the nucleation barriers (B) can be treated as a random variable with a Gaussian probability distribution ($p(B)$). In this case, we will assume the distribution has a mean barrier ($\bar{B}$) and a variance (σ^2). The barrier probability distribution is:

$$[12] \quad p(B) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\left(B - \bar{B}\right)^2}{2\sigma^2}\right)$$

and, assuming the prefactor A remains the same from impurity to impurity, the nucleation rate is:

$$[13] \quad rate = A\int dB \, p(B)\exp\left(\frac{-B}{S^2}\right) = A\exp\left(\frac{-\bar{B}}{S^2} + \frac{\sigma^2}{2S^4}\right)$$

The effect of the impurities is that the rate depends less strongly on the degree of supersaturation. Applying equation [7] to equation [13] yields the average nucleus size for this heterogeneous system:

$$[14] \quad \langle n* \rangle = \frac{2\bar{B}}{S^3} - \frac{2\sigma^2}{S^5}$$

Comparing equation [14] to equation [11], one can see that the average nucleus size for a heterogeneous process will have less of a dependence on the degree of

158

supersaturation than a process with a single barrier. This is due to the compensatory effect of the variance term in equation [14]. This effect is illustrated in Figure 4.1 where the dependence of the observed average nucleus size ($<n^*>$) is plotted against the degree of supersaturation for a homogeneous process and a heterogeneous process.
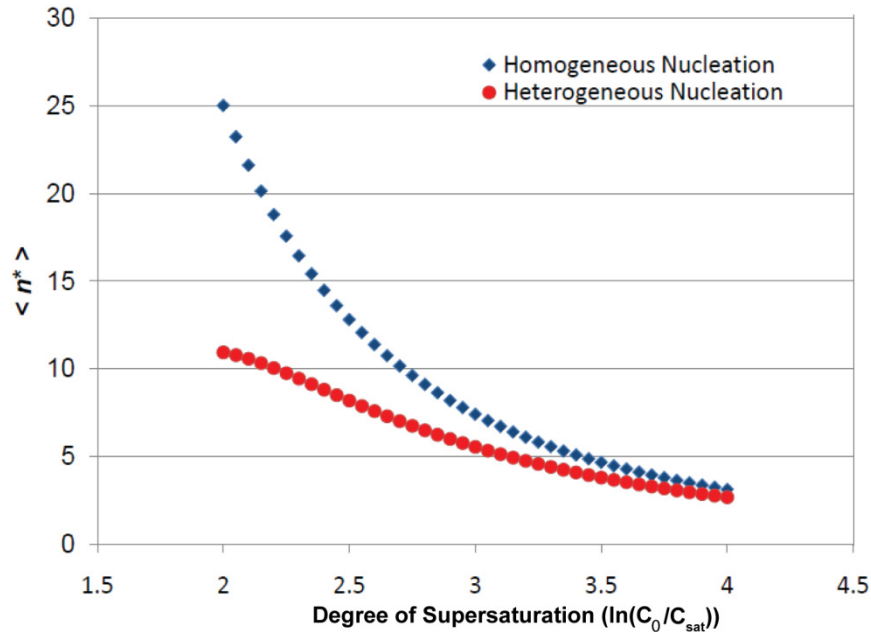
Figure 4.1 A comparison of the average nucleus size ($<n^*>$) for homogeneous nucleation with a single barrier (B =100) and heterogeneous nucleation with a Gaussian distribution of barriers with a mean barrier ($\bar{B}$ =100) and a variance (σ^2 =225).

4.2.3 Connecting CNT with the pre-equilibrated nucleus model[1]

The model used by Wetzel and coworkers (*12*) to assess the nucleation behavior of polyglutamine was already discussed in detail[1], it is useful to point out a couple of the

---

[1] This model was described in Scheme 4 of Chapter 1, Section 1.5.3.2.

equations in this model which provide a direct connection to CNT. In their model, the

initial rate of monomer loss (Δ) is assumed to be constant and is given by:

$$[15] \quad \Delta = \frac{1}{2} C_0 k^+ J^* c^* t^2$$

where C_0 is the initial protein concentration, k^+ is the rate constant for elongation of

growing fibrils, J^* is the rate constant for elongation of the nucleus, t is time, and c^* is

the concentration of nuclei. It can be seen that the nucleation rate can be solved as:

$$[16] \quad J^* c^* = \frac{2\Delta}{C_0 k^+ t^2}$$

Equation [15] is analogous to the rate given by equations [1] or [8] for CNT. Assuming

the reaction is accurately described as homogeneous nucleation, one can calculate the

nucleation rate from knowledge of k^+, C_0, and $\frac{\Delta}{t^2}$. Wetzel and coworkers have measured

the fibril elongation rate constant, k^+, to be $\sim 10^4$ M^{-1}s^{-1} for both KKQ$_{23}$KK and

KKQ$_{37}$KK. (1, 5) This is the value we use in our analysis.

In this Chapter, we use equations [15] and [16] to calculate the nucleation rates

for KKQ$_{18}$KK, KKQ$_{23}$KK, and KKQ$_{37}$KK from previously published plots of $\frac{\Delta}{t^2}$ versus

C_0 from the work of Kar et al. (5) We then apply CNT to the rates and calculate the

expected nucleus size for this reaction, assuming that homogeneous nucleation applies.

This requires knowledge of the degree of supersaturation, S, defined as $S = \ln(C_0/C_{sat})$.

Under the same solution conditions used for the kinetics experiments, Wetzel and

coworkers have measured C_{sat} to be 3 μM for KKQ$_{23}$KK. (5) They have also measured

the saturation concentrations for KKQ$_{15}$KK and KKQ$_{40}$KK to be approximately 30 μM

and 70 nM, respectively. (*2*) Based on these numbers, we estimated C_{sat} to be 15 µM and 80 nM for $KKQ_{18}KK$ and $KKQ_{37}KK$, respectively.

4.3 Results

Applying equations [15] and [16] to plots of $\frac{\Delta}{t^2}$ versus C_0 from the work of Kar et al. (*5*), led to the calculations of nucleation rates shown in Figure 4.2 plotted against the degree of supersaturation. These rates were fit to equation [8] which yielded values for A and B as shown in equation [8] and equations [9] and [10]. The data were fit using the Levenberg-Marquadt algorithm (*15, 16*) to minimize the sum-of-the-square of the residuals between the data and the calculated fit. The calculated values of A and B for each construct are shown in Table 4.1. The R^2 measured goodness-of-fit was greater than 0.94 for all fits.

Figure 4.2 The nucleation rates measured by Kar et al. (*5*) for $KKQ_{18}KK$, $KKQ_{23}KK$, and $KKQ_{37}KK$ are plotted against the degree of supersaturation. The data are fit to equation [8], which describes the nucleation rate for a homogeneous nucleation process according to CNT. The R^2 measured goodness-of-fit was >0.94 for all fits. The parameters *A* and *B* resulting from the fits with equation [8] are shown in Table 4.1.

	$KKQ_{18}KK$	$KKQ_{23}KK$	$KKQ_{37}KK$
A ($cm^{-3}s^{-1}$)	5814	8446	6.9
B	163.4	167.8	112.3

Table 4.1 The parameters *A* and *B* were determined by fitting equation [8] to the nucleation rates measured by Kar et al. (*5*) for $KKQ_{18}KK$, $KKQ_{23}KK$, and $KKQ_{37}KK$ plotted against the degree of supersaturation (Figure 4.2).

Figure 4.2 suggests that KKQ$_{18}$KK and KKQ$_{23}$KK appear to have a different aggregation behavior than KKQ$_{37}$KK, and this is recapitulated in Table 4.1. It is important to note that the experiments for KKQ$_{37}$KK were performed at higher degrees of supersaturation compared to KKQ$_{18}$KK and KKQ$_{23}$KK.

As shown in equation [7], the rate of change of ln(*rate*(*S*)) as a function of *S* yields an estimate of the nucleus size as a function of *S*. A plot of the log of the nucleation rates measured by Kar et al. (*5*) for KKQ$_{18}$KK, KKQ$_{23}$KK, and KKQ$_{37}$KK plotted against the degree of supersaturation are shown in Figure 4.3.
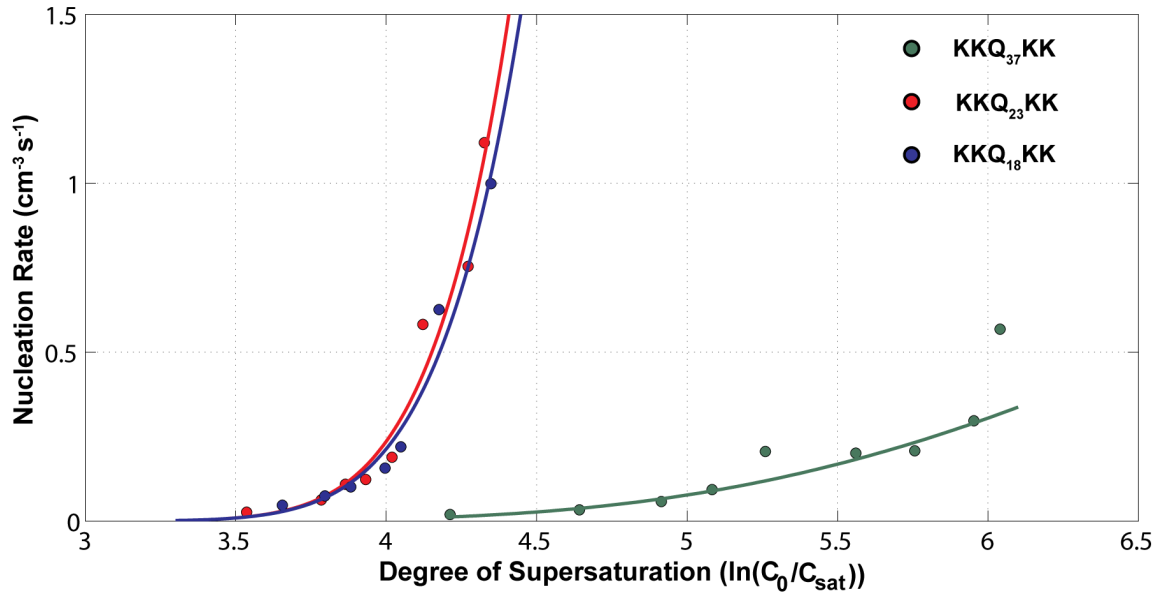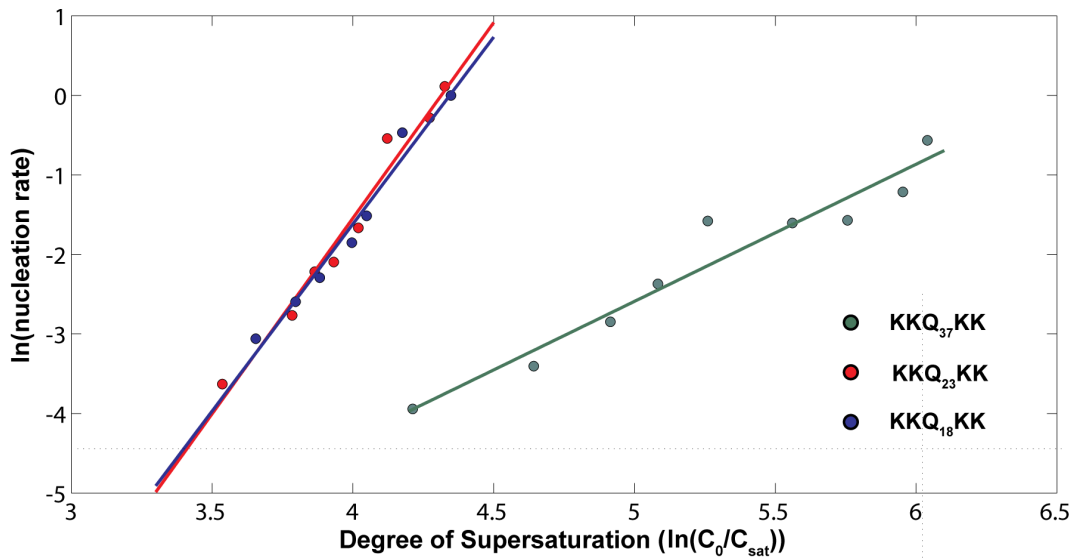

Figure 4.3   The logarithm of the nucleation rates measured by Kar et al. (*5*) for KKQ$_{18}$KK, KKQ$_{23}$KK, and KKQ$_{37}$KK are plotted against the degree of supersaturation. The data are well fit by a line ($R^2 > 0.96$ for all fits). This suggests, from equation [7],

that the nucleus size is constant and is given by the slope of the line of best fit. The slopes are 4.7 for $KKQ_{18}KK$, 4.9 for $KKQ_{23}KK$, and 1.7 for $KKQ_{37}KK$.

The results from Figure 4.3 suggest that the nucleus size is constant over all degrees of supersaturation investigated. This is inconsistent with equation [11], which, for a homogeneous nucleation, suggests that the nucleus size should vary (*i.e* decrease) with the degree of supersaturation. Importantly, for a homogeneous nucleation process, the prediction of the nucleus size from equation [7] and equation [11] will be consistent. A plot of the predicted nucleus size predicted from equation [7] and the predicted nucleus size from equation [11] are plotted against the degree of supersaturation for the constructs $KKQ_{18}KK$, for $KKQ_{23}KK$, and $KKQ_{37}KK$ in Figure 4.4. In Figure 4.4, the nucleus size predicted from equation [7] is represented as a horizontal line because the value is constant over all concentrations investigated, and the nucleus size predicted from equation [11] is represented as filled circles.

164
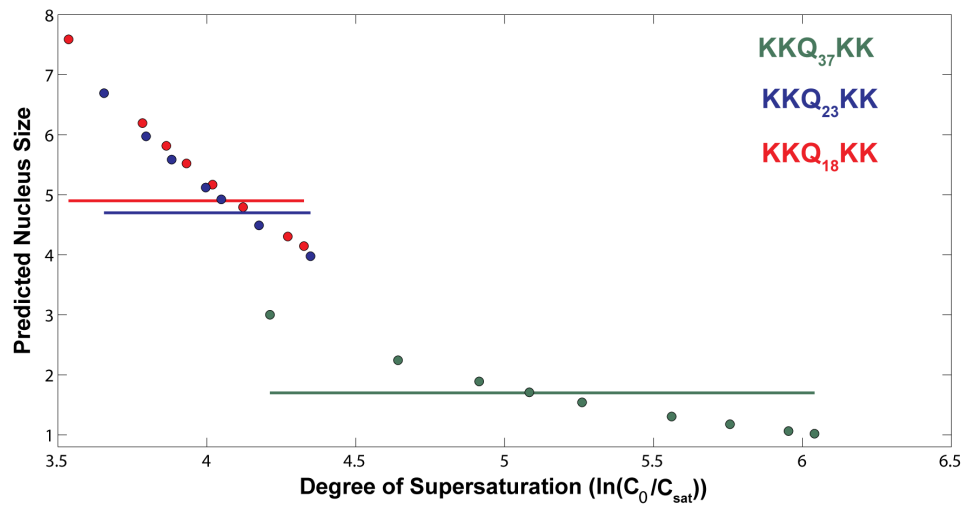
Figure 4.4 The predicted nucleus size (based on the experiments of Kar et al. (*5*) for the aggregation of polyglutamine constructs $KKQ_{18}KK$, $KKQ_{23}KK$, and $KKQ_{37}KK$) is plotted against the degree of supersaturation. The nucleus size predicted from equation [7] is represented as a straight line, and the nucleus size predicted from equation [11] is represented as filled circles. If homogeneous nucleation applies, the lines should overlap with the filled circles, but they do not.

The data presented in Figures 4.3 and 4.4 suggest that the aggregation process is not well characterized as a homogeneous nucleation process.

4.4 Discussion

The analysis presented in this chapter compares the predictions for homogeneous nucleation based on classical nucleation theory (CNT) to those made by a so-called thermodynamic nucleus model (*12*) for describing the aggregation of polyglutamine.

CNT provides clear diagnostic criteria for establishing whether or not a process can be classified as homogeneous nucleation. According to these criteria, $n*$ in a homogeneous nucleation process should depend on the degree of supersaturation (equation [11]). This means that analysis of the nucleation rate of a reaction using either equation [7] or equation [11] should yield the same estimate of the nucleus size for a given degree of supersaturation.

Analysis of the rates of nucleation for the aggregation reactions presented in Kar et al. (*5*) for the polyglutamine constructs $KKQ_{18}KK$, $KKQ_{23}KK$, and $KKQ_{37}KK$ suggests that the nucleus size for a given construct is constant over all concentration ranges measured (Figure 4.4, solid lines). For a homogeneously nucleated process, the nucleus size should show an inverse dependence on the cube of the degree of supersaturation (Figure 4.4, filled circles). The data in Figure 4.4 suggest that the process is not well characterized as a homogeneous nucleation.

Based on the phase diagrams presented in Chapter 3, we hypothesize that the aggregation reaction will proceed through heterogeneous nucleation. The AFM data in Chapter 3 support this hypothesis. In addition, Vitalis and Pappu (*17*) have shown that analysis of a heterogeneous process with the thermodynamic nucleus model (*12*) used by Wetzel and coworkers will lead to fractional estimates for the nucleus size. The nucleus sizes estimated for $KKQ_{18}KK$, $KKQ_{23}KK$, and $KKQ_{37}KK$ using the pre-equilibrated nucleus model were 3.7, 3.9, and 0.7, respectively. (*5*) Thus, the analysis presented here

is consistent with the work of Vitalis and Pappu (*17*) and suggests that the aggregation of polyglutamine constructs should not be characterized as homogeneous nucleation.

It is important to reiterate that a plot of the logarithm of the nucleation rate against the degree of supersaturation (or even the logarithm of the starting concentration as was done in the work of Wetzel and co-workers) which produces a straight line is unlikely to reflect homogeneous nucleation, which is in direct contrast to the assumptions made in the model used by Wetzel and co-workers (*12*). To accurately diagnose a process as a homogeneously nucleation, one should perform the experiments over a wide range of degrees of supersaturation. The model used by Wetzel and co-workers (*12*) is designed to be used only over a very small range of degrees of supersaturation. This model, as applied by Kar et al. (*5*) will never show a dependence of the nucleus size on the degree of supersaturation because the authors explicitly assume that the nucleation rate is constant over all concentrations measured (see equations [15] and [16]). According to CNT, this assumption is not valid.

4.5 Summary

We have shown in this chapter that knowledge of the phase diagram provides a direct link to CNT through knowledge of C_{sat}, which, in turn, defines the degree of supersaturation (S) for a given starting concentration of protein (equation [3]). The rate

of nucleation is proportional to $\exp\left(\dfrac{-1}{S^2}\right)$ in CNT, and this is true for both homogeneous and heterogeneous nucleation.

We have also shown that the proposal that aggregation of KKQ_NKK is a homogeneous nucleation and growth process with a critical nucleus size of one molecule for $N > 26$ is not consistent with CNT. A re-analysis of the same data with CNT suggests that the aggregation process is better characterized as heterogeneous nucleation, and this is consistent with results from Vitalis and Pappu (*17*), the interpretations of the phase diagrams, and the AFM images presented in Chapter 3.

4.6 References

1. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction, *Proc Natl Acad Sci U S A 102*, 15400-15405.
2. Chen, S., Berthelier, V., Hamilton, J. B., O'Nuallain, B., and Wetzel, R. (2002) Amyloid-like features of polyglutamine aggregates and their assembly kinetics, *Biochemistry 41*, 7391-7399.
3. Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001) Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity, *J. Mol. Biol. 311*, 173-182.
4. Chen, S., Ferrone, F. A., and Wetzel, R. (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation, *Proc Natl Acad Sci U S A 99*, 11884-11889.
5. Kar, K., Jayaraman, M., Sahoo, B., Kodali, R., and Wetzel, R. (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent, *Nat Struct Mol Biol 18*, 328-336.
6. Lee, C. C., Walters, R. H., and Murphy, R. M. (2007) Reconsidering the mechanism of polyglutamine peptide aggregation, *Biochemistry 46*, 12810-12820.

7. Rubinstein, M., and Colby, R. H. (2003) *Polymer Physics*, Oxford University Press, Oxford and New York.
8. Sear, R. P. (2007) Nucleation: theory and applications to protein solutions and colloidal suspensions, *Journal of Physics: Condensed Matter 19*, 033101.
9. Oxtoby, D. W., and Kashchiev, D. (1994) *J. Chem. Phys. 100*, 7665.
10. Oxtoby, D. W. (1992) *J. Phys.: Condens. Matter 4*, 7627.
11. Debenedetti, P. G. (1996) *Metastable Liquids*.
12. Ferrone, F. (1999) Analysis of protein aggregation kinetics, *Methods Enzymol 309*, 256-274.
13. Turnbull, D., and Vonnegut, B. (1952) *Ind. Eng. Chem. 44*, 1292.
14. Turnbull, D. (1950) *J. Chem. Phys. 18*, 198.
15. Levenburg, K. (1944) A Method for the Solution of Certain Non-Linear Problems in Least Squares, *The Quarterly of Applied Mathematics 2*, 164-168.
16. Marquardt, D. W. (1963) An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics 11*, 431-441.
17. Vitalis, A., and Pappu, R. V. (2011) Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides, *Biophys Chem*.

Chapter 5: Amyloid seeds formed by cellular uptake, concentration, and aggregation of the amyloid-beta peptide[1]

5.1 Preamble

Chapter 5 is a departure from previous chapters in two major ways: 1) aggregation is studied in cells as opposed to test tubes, and 2) the system of interest is amyloid-beta (Aβ), the major component of extracellular plaques in Alzheimer's disease patients. As we have shown in previous chapters, the saturation concentration, at least for polyglutamine lies in the low micromolar range in 50 mM phosphate buffer. Others have found that the saturation concentration of Aβ in aqueous milieus also lies in this concentration regime. (*1, 2*) However, it is known that the *in vivo* concentration of Aβ in the cerebrospinal fluid lies in the low nanomolar range, which, based on the determined value for the saturation concentration, suggests that aggregation or at least precipitation should be thermodynamically unfavorable. (*3-7*) Aggregates of Aβ are one of the major pathological hallmarks of Alzheimer's disease, and there must be some mechanism *in vivo* by which the saturation concentration barrier is crossed. This can be done in one of

---

[1] This work has been published with S.L. Crick and X. Hu contributing equally*:

Hu, X.*, Crick, S. L.*, Bu, G., Frieden, C., Pappu, R. V., and Lee, J. M. (2009) Amyloid seeds formed by cellular uptake, concentration, and aggregation of the amyloid-beta peptide, *Proceedings of the National Academy of Sciences of the United States of America 106*, 20324-20329.

two ways: by locally concentrating Aβ or by lowering the saturation concentration either through modifications to the protein or changes in solution conditions. In this chapter we will focus on an *in vivo* mechanism leading to effective concentrations and solution conditions under which Aβ aggregation is favorable.

5.2 Introduction

Alzheimer's disease (AD), the most common form of dementia in Western countries, involves progressive accumulation of amyloid deposits, neuronal loss, cognitive decline, and eventual death. Senile plaques, a key pathological feature of this disease, are composed primarily of the amyloid-beta (Aβ) peptide, and are found throughout the brain (*8*). Aβ (ranging in length from 39-42 amino acids) is derived from the proteolytic cleavage of an endogenous transmembrane protein known as the amyloid precursor protein (APP). The most common Aβ peptide found in senile plaques is the 42-residue peptide (Aβ$_{1-42}$) (*9*), which also shows the strongest propensity for spontaneous aggregation in solution (*10*). It is widely believed that the aggregation and accumulation of this peptide is involved in disease pathogenesis.

Aβ is produced by neurons and secreted into the brain extracellular space where it is normally found in a soluble state (*11*). A variety of physiological processes, including those associated with neuronal activity, are related to Aβ synthesis and release into the extracellular space (*12-14*). Under normal physiological conditions and in AD patients, the concentration of Aβ in brain extracellular fluid (interstitial fluid, ISF and

cerebrospinal fluid, CSF) is low (10^{-10} M – 10^{-9} M) (*3-7*). This is important because *in vitro* studies suggest that the saturation concentration (C_{sat}) for spontaneous aggregation of Aβ is in the μM range (*1, 2*). Therefore, Aβ concentrations *in vivo* would have to increase by 3-4 orders of magnitude or the C_{sat} would have to decrease by at least 3-4 orders of magnitude for phase separation to be feasible. In order to span this large concentration gap, several potential mechanisms have been proposed. The local concentration of Aβ could be increased through membrane association (*15-17*), macromolecular crowding (*18*), or specific interactions with protein complexes or chaperones (*19-22*).

Alternatively, one could imagine that the C_{sat} of Aβ might be lowered. We have seen in Chapter 3 that the free energy of mixing (and, hence, solubility) is a function of both chain-chain and chain-solvent interactions. Therefore the C_{sat} of Aβ could be decreased either by altering the properties of Aβ or by altering the solution conditions or both. Researchers have shown that covalent modifications, such as linkage of two Aβ molecules through a tyrosine-tyrosine bond (*23*), site-specific addition of cholesterol to a lysine sidechain (*24*), or N-terminal truncation of Aβ containing a pyroglutamate (*25*) can lower the C_{sat} of Aβ, although the degree of lowering is not well-quantified. It has also been shown that the C_{sat} of Aβ is known to be sensitive to solution pH, with the minimum C_{sat} occurring near the peptide's isoelectric point (pI~5.3 for Aβ$_{1-40}$). (*26*) It has also been shown the C_{sat} of Aβ is sensitive to the presence of certain metal ions (*27*).

One might imagine that all of these factors including: local concentration increases, covalent peptide modifications, and fluctuations in solution conditions likely play a role in the aggregation of Aβ in the human brain.

It is widely believed that Aβ aggregation is best-characterized as a nucleation-elongation process. (*1, 26, 28-30*) Assuming this is the case, the barrier to aggregation is the formation of the nucleus. We have already seen (see Chapter 4) that the probability of forming a nucleus is a function of the degree supersaturation of a solution. Therefore, any process which lowers the peptide's C_{sat} or increases the peptide's concentration will lead to a higher probability of forming the nucleus.

The lysosome is an attractive place for Aβ to form the nucleus *in vivo.* These acidic vesicles have a pH close to the peptide's isoelectric point (which means that the C_{sat} is at a minimum), and the peptide's C_{eff} (the effective concentration of Aβ in a confined vesicular environment) might be higher than C_{sat}. Indeed, researchers proposed that lysosomes might be a perfect environment for the initiation of protein aggregation leading to AD. (*31-33*) In this chapter, we explore the possibility that low concentrations of soluble Aβ can be taken up by neurons and concentrated into acidic vesicles where the combination of low pH and high local concentration yield favorable conditions for the spontaneous aggregation of Aβ.

5.2 Materials and Methods

5.2.1 Reagents

$A\beta_{1-42}$ and $A\beta_{1-40}$ were purchased in their lyophilized form from American Peptide (Sunnyvale, CA). Fluorescein isothiocyanate (FITC) labeled Aβ, both FITC-$A\beta_{1-42}$ and FITC-scrambled-$A\beta_{1-42}$ (where scrambled indicates that protein has a random sequence with the same overall amino acid content as the wild-type protein) were purchased in their lyophilized form from rPeptide (Bogart, GA). Tetramethylrhodamine (TMR), TMR-scrambled-Aβ, and TMR-$A\beta_{1-42}$ were purchased in their lyophilized form from Anaspec (Fremont, CA). Fluorescein, trifluoroacetic acid (TFA), and 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) were purchased from Sigma (St. Louis, MO). LysoTracker® Red DND-26 and LysoTracker® Green DND-26 were purchased from Molecular Probes (Eugene, OR). The 6E10 antibody was purchased from Sigma and the 3D6 antibody was a gift from Eli Lilly (Indianapolis, IN).

5.2.2 Aβ Preparation

$A\beta_{1-42}$ and $A\beta_{1-40}$ were prepared by dissolving the dry peptide (1 mg) in a Wheaton glass vial with neat TFA (1 ml). This solution was dried under a gentle nitrogen stream and then resuspended in HFIP (1 ml). This solution was again dried under a gentle nitrogen stream leaving a peptide film on the glass vial. The peptide was then dissolved in neat DMSO at a concentration of 200 µM and stored at $-20°$ C until use.

5.2.3 Cell Culture and Cellular Uptake

Neuoroblastoma SHSY5Y, murine cortical neurons and HEK293 cells were grown in Dulbeco's modified eagle medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (FBS). Cortical neurons derived from embryonic day 15.5 mice were grown in BME medium (with 5% Horse serum, 5% FBS) and treated with AraC (Cytosine β-D-arabinofuranoside hydrochloride, 4ug/ml) at 1 day. After 7 days, 250 nM FITC-Aβ_{1-42} was added to culture media. As controls, at 7 days *in vitro*, 250 nM free fluorescein, free TMR, FITC- Aβ_{1-40}, and FITC-scrambled-Aβ_{1-42} were added to SHSY5Y cell cultures. Cells were imaged using a Confocor II LSM system (Carl Zeiss-Evotec) at 24 hours after addition of the fluorescent species.

For dose and time dependence experiments, SHSY5Y cells were plated in 8-well Lab-Tek chamber slides (Nunc). For time dependence experiments, 24 hours after plating the cells, TMR-Aβ_{1-42} was diluted in DMEM-10% FBS and added to the cell culture at a concentration of 250 nM. For the dose dependence experiments, TMR-Aβ_{1-42} was added to make the final concentration. Images were taken 24 hours after addition of the TMR-Aβ_{1-42} using a Zeiss confocal laser scanning microscope.

5.2.4 Costaining of Aβ and LysoTracker

SHSY5Y cells were treated with 100 nM fluorescein Aβ_{1-42} for 24 hours and then LysoTracker Red or Green (a dye specific for acidic vesicles) was added to the medium

to a final concentration of 50 nM. After incubation at 37°C for 30 min, images were obtained using a Zeiss confocal laser scanning microscope.

5.2.5 Confocal Microscopy

All measurements were performed on a Confocor II LSM system (Carl Zeiss-Evotec) with a 40X water immersion objective. SHSY5Y cells were grown in Labtek 8-well chambers. After being incubated with 25 nM TMR-$A\beta_{1-42}$ for 24 hours, the cells were imaged. The samples were excited at 543 nm with a helium-neon laser, and emissions were collected from 560 to 615 nm using a bandpass filter. The correlation between pixel intensity for different excitation laser powers and concentration was calibrated using known concentrations of free dye (TMR). Images of the cells were taken under the same microscope conditions used during the calibration. Six images were analyzed at 5 different excitation laser powers. The images were collected at random regions of the cell culture dish. In each image, 4-5 fluorescent points, of size consistent with lysosomes (1-3 μM in diameter), were selected inside different cells. For each point, four adjacent pixels centered about the most intense region of the vesicle were analyzed. The mean brightness was calculated and compared to the calibration curves to estimate the intravesicular $A\beta$ concentration.

5.2.6 Cell Death Assay

Lactate Dehydrogenase (LDH) Activity released into culture media by dying cells was assessed as an index of cell death. SHSY5Y cells were grown in varying

concentrations of $A\beta_{1-42}$ (0-1000 nM as indicated) for 5 days. 25 µl of medium from the culture was mixed with 125 µl LDH buffer (33 mM KH2PO4, 66 mM K2HPO4, pH7.4) and 100 µl NADH solution (0.03% NADH in LDH buffer, freshly made before assay) for 10 min. Then 25 µl pyruvate solution was added immediately before reading at 340 nm. Cells sonicated in 1% Triton were used to define "full kill."

5.2.7 Immunoblot

SHSY5Y cells were washed twice with cold PBS and then detached by 0.05% trypsin/0.02% EDTA after incubation with 1 µM $A\beta_{1-42}$ for 3 days. The cell pellet was resuspended in cold Hanks' balanced salt solution (HBSS buffer), followed by sonication at level 3 for 10 times. Cell lysates were centrifuged at 3000xG for 5 minutes. The supernatant was loaded onto a 16.5% Tris-Tricine gel, transferred to a polyvinylidene difluoride (PVDF) membrane and probed with monoclonal antibody 6E10.

5.2.8 Agarose Gel electrophoresis

Agarose (1.5% w/v) was melted in running buffer (20 mM Tris, 200 mM Glycine) without SDS. While stirring the melted agarose, 10% (w/v) SDS solution was added (drop by drop to avoid local solidification of the agarose) to a 0.1% final SDS concentration. Cell extract was incubated for 7 min in sample buffer (60 mM Tris-HCl (pH 6.8), 5% glycerol, 2% SDS, 0.05% bromphenol blue) at room temperature, resolved

in a horizontal 1.5% agarose gel in a standard Tris/glycine/SDS buffer, transferred electrophoretically to a polyvinylidene difluoride membrane, and probed with 6E10.

5.2.9 *In vitro* seeding

SHSY5Y cells were washed twice with cold PBS and then detached by 0.05% trypsin/0.02% EDTA. After centrifugation, the cell pellet was resuspended in 2% SDS in PBS, and sonicated (Fisher Scientific) at level 10 using 1 sec pulses x 20. Cell lysates were diluted 1:40 in PBS with 0.02% sodium azide and loaded into Labtek 8-well chambers, to a final SDS concentration of 0.05%. 100 nM TMR-Aβ_{1-42} was then added, and after incubation at 37°C for 48 hours, images were acquired by confocal microscopy.

5.2.10 Thioflavin-S staining

Culture wells were washed twice with PBS, followed by incubating with 0.025% Thioflavin-S solution (0.025 g Thioflavin-S dissolved in 100 ml 50% ethanol) for 5 minutes and then rinsed with 50% ethanol twice. After washing with MilliQ water twice, the wells were cover slipped using mounting medium for the imaging.

5.3 Results

5.3.1 Uptake of human Aβ into late endosomes/lysosomes of neural cells

In separate experiments, 250 nM of $A\beta_{1-42}$ labeled with fluorescein isothiocyanate ($FITC-A\beta_{1-42}$) was incubated with a variety of cell types for 24 hours and then imaged using confocal microscopy. Uptake of $FITC-A\beta_{1-42}$ was observed in vesicles of SHSY5Y neuroblastoma cells (Figure 5.1A) and murine cortical neurons (Figure 5.1B), while none was observed in hamster embryonic kidney (HEK293) cells (Figure 5.1C). These observations, consistent with previously published findings (*32*), suggest cellular specificity in the uptake of $FITC-A\beta_{1-42}$. To examine the sequence specificity for uptake of Aβ, SHSY5Y cells were incubated with 250 nM of $FITC-A\beta_{1-40}$ or FITC-scrambled-$A\beta_{1-42}$ (FITC-labeled peptide that shares amino acid composition with Aβ but has its sequence randomized). Vesicular uptake was observed with $FITC-A\beta_{1-40}$ (Figure 5.1F) but not FITC-scrambled-$A\beta_{1-42}$ (Figure 5.1E), suggesting a sequence-specific uptake mechanism. No uptake was observed, however, when SHSY5Y cells were incubated with the fluorescein label alone (Figure 5.1D). Because of the ease of growth and manipulation (relative to primary neuronal cultures), SHSY5Y cells were used to perform further studies.

179
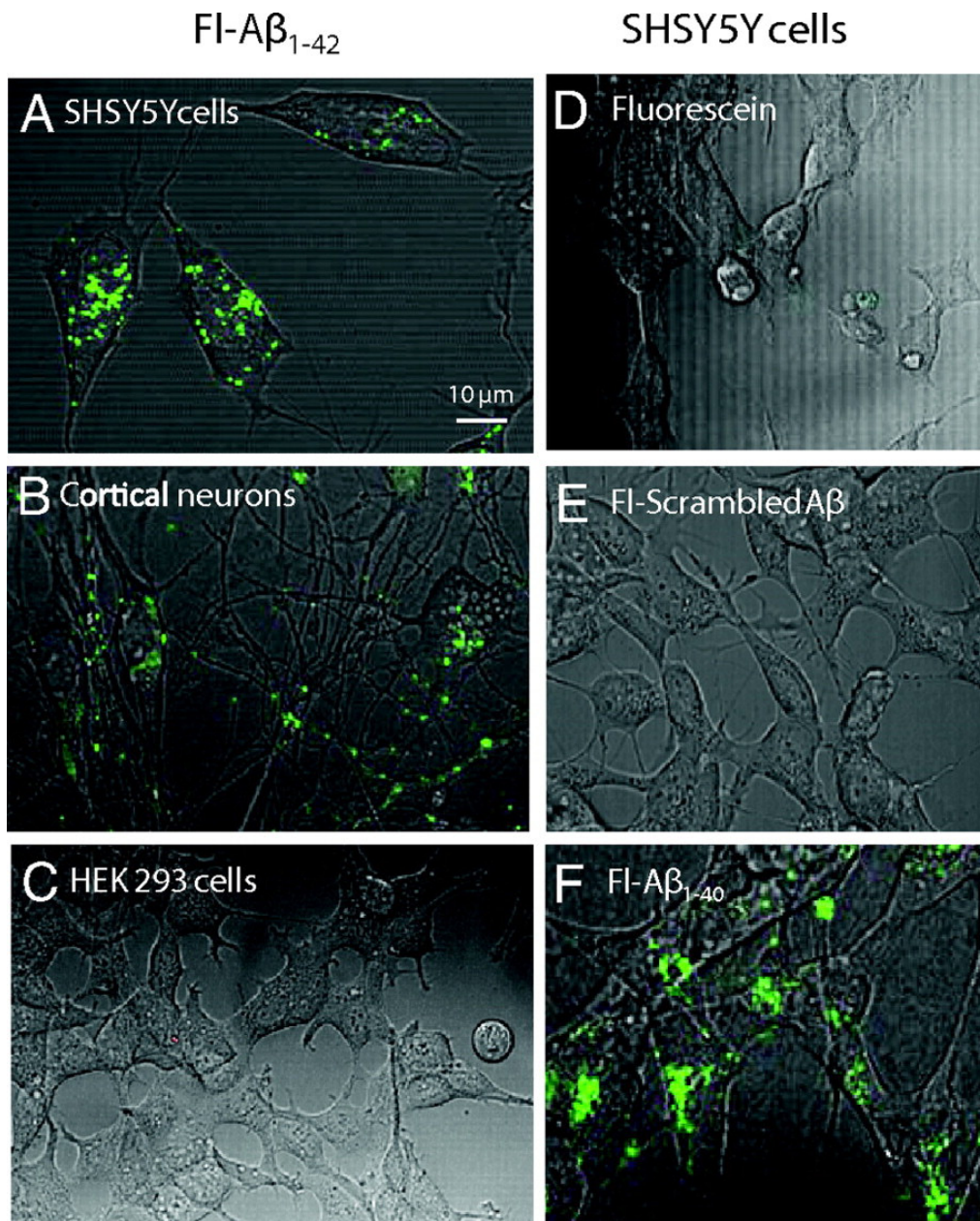
Figure 5.1 Neural cell uptake of FITC-Aβ. A) SHSY5Y human neuroblastoma cells, B) HEK293 cells, and C) primary murine cortical neurons were cultured in the presence of 250 nM FITC-labeled synthetic human Aβ$_{1-42}$ for 24 hours and then imaged with confocal microscopy. Vesicular uptake was observed only in the neural cells (SHY5Y and cortical neurons). (D) SHSY5Y cells were incubated with 250 nM fluorescein alone, (E) FITC-

scrambled Aβ$_{1-42}$, or (F) FITC-Aβ$_{1-40}$ for 24 hours. Vesicular uptake was observed with FITC-Aβ$_{1-42}$ and Aβ$_{1-40}$, but not with FITC-scrambled-Aβ$_{1-42}$ or fluorescein alone.

In order to identify the vesicles into which Aβ accumulated, SHSY5Y cells loaded with 250 nM FITC-Aβ$_{1-42}$ for 24 hours were treated with LysoTracker (a pH-sensitive, membrane-permeable dye that stains acidic vesicles including late endosomes and lysosomes) 30 minutes prior to imaging. All vesicles with FITC-Aβ$_{1-42}$ co-stained with LysoTracker and appeared to be a large subset of LysoTracker-positive vesicles (Figure 5.2), suggesting that that both FITC-Aβ$_{1-42}$ and FITC-Aβ$_{1-40}$ are taken up and trafficked to acidic vesicles (e.g. late endosomes or lysosomally-derived vesicles). Because the fluorescence of fluorescein decreases at low pH (*34*), it is not an optimal fluorophore for measuring the extent of uptake into acidic vesicles. For more quantitative studies on the uptake of labeled Aβ$_{1-42}$ by SHSY5Y cells, tetramethylrhodamine (TMR) was used as the fluorophore because its fluorescence is relatively insensitive to pH. (*35*)
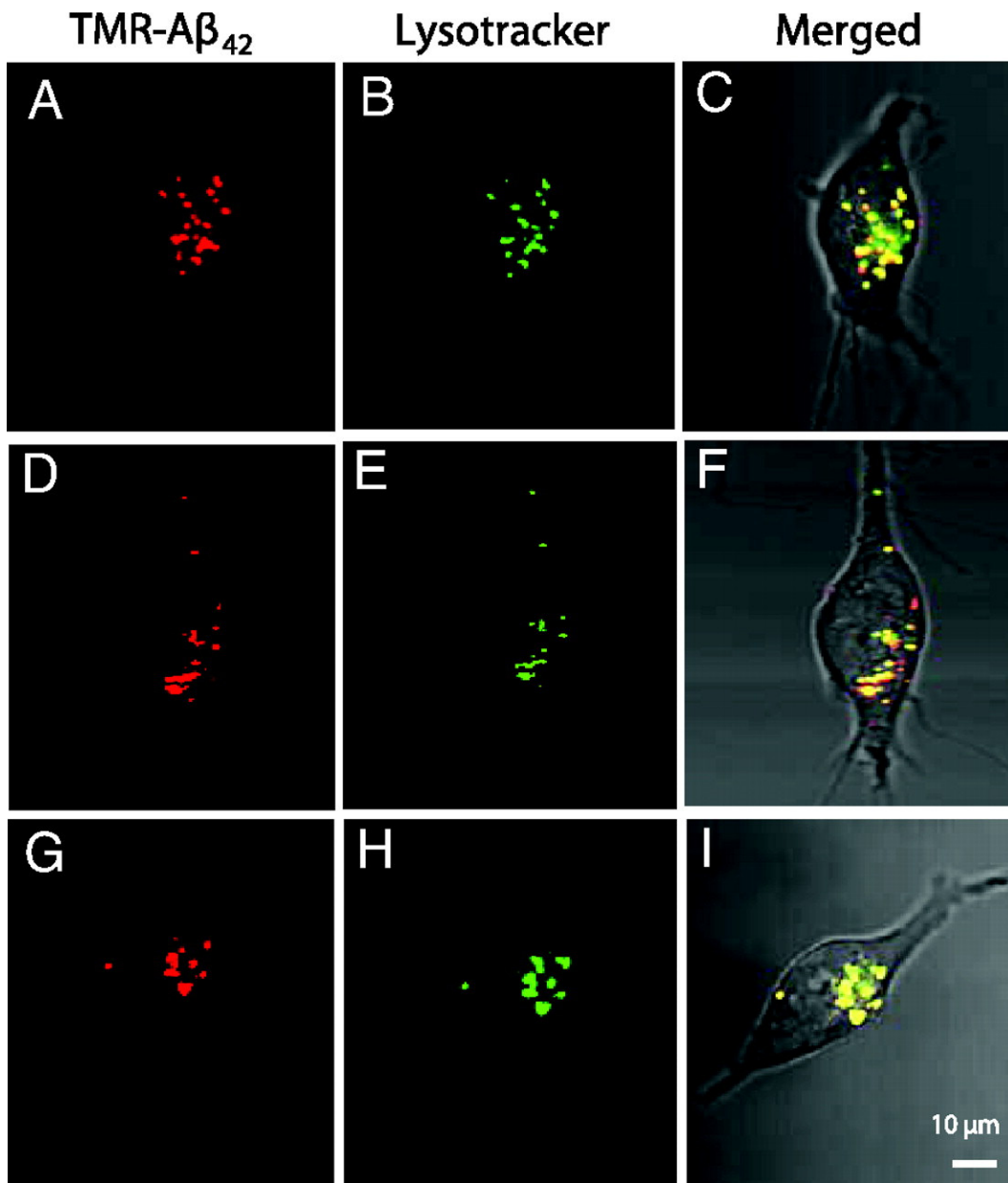
Figure 5.2 Intracellular co-localization of Aβ$_{1-42}$ with LysoTracker. SHSY5Y cells were grown in the presence of 250 nM FITC-Aβ$_{1-42}$ for 24 hours and imaged 30 minutes after 50 nM LysoTracker was added to the culture medium. (A,D,G) TMR- Aβ$_{1-42}$ was detected in vesicles that co-stained with LysoTracker (B,E,H); merged fluorescent images

with phase contrast image (C,F,I) demonstrate co-localization of Aβ and LysoTracker-stained vesicles.
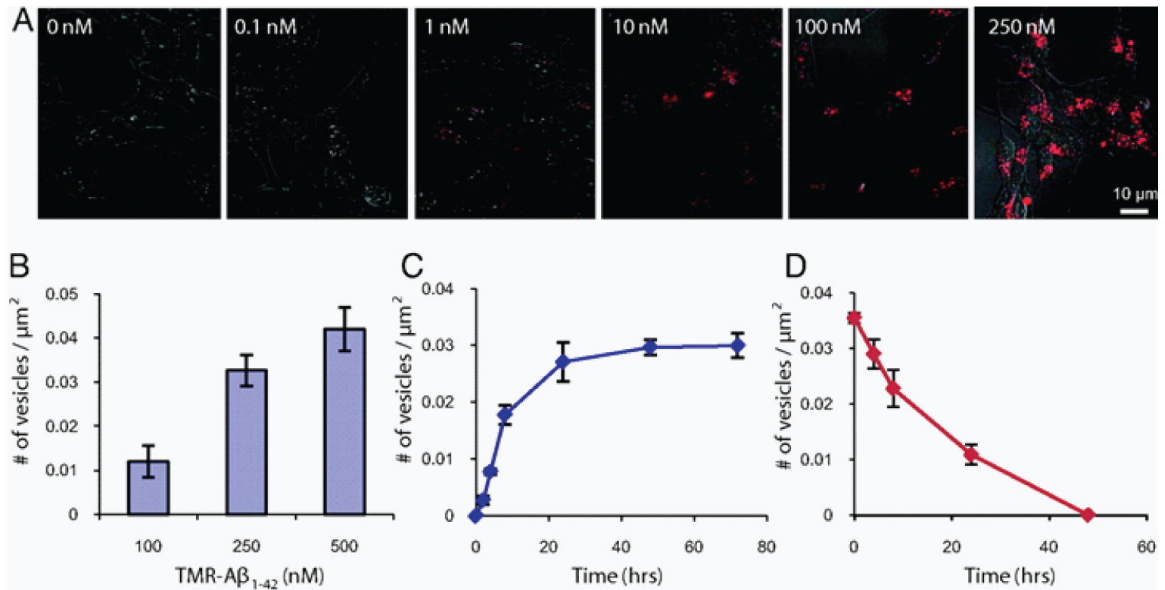
Figure 5.3 Dose- and time-dependence of vesicular uptake of TMR-Aβ$_{1-42}$. SHSY5Y cells were grown in varying concentrations of TMR-labeled Aβ$_{1-42}$ (1–250 nM as indicated (*A*), then imaged using confocal microscopy after 24 hours. Fluorescent vesicles were quantified, and demonstrated dose-dependent uptake (*B*). SHSY5Y cells were grown in 250 nM TMR-Aβ$_{1-42}$, imaged at varying times (0–72 hours) thereafter, and density of fluorescent vesicles was quantified (*C*). SHSY5Y cells were grown in 250 nM TMR-Aβ$_{1-42}$ for 24 h, TMR-Aβ$_{1-42}$ was washed out of the medium, imaged at various times thereafter, and density of fluorescent vesicles was quantified (*D*). After washout, the number of fluorescent vesicles decreased with time, and disappeared by 48 hours. Error bars represent the standard error of the mean from three independent experiments.

Uptake of TMR-Aβ_{1-42} was similar to that of FITC-Aβ_{1-42}, and was both dose- (Figures 5.3A and 5.3B) and time-dependent (Figure 5.3C), indicating that uptake is not affected by the choice of fluorophore. Extracellular concentrations as low as 1 nM were sufficient to produce vesicular uptake within 24 hours of incubation observable under confocal microscopy (Figures 5.3A and 5.3B). Furthermore, fluorescent vesicles were first apparent within 4 hours after the addition of 250 nM TMR-Aβ_{1-42} to cell cultures (Figure 5.3C). Incubation with TMR alone did not result in any intracellular fluorescence throughout the concentration range examined. Furthermore, cell death was not detectable at all concentrations of TMR- Aβ_{1-42} examined (Figure 5.4).
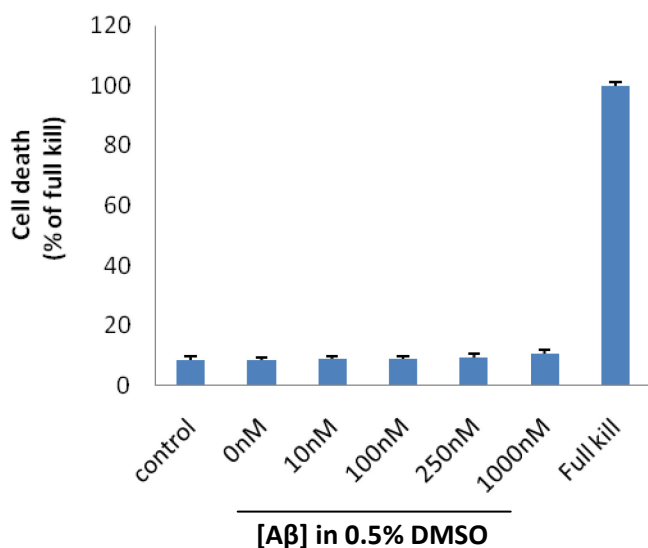
Figure 5.4 Cell death after Aβ uptake. SHSY5Y cells were grown in the presence of varying concentrations of Aβ_{1-42} (0-1000 nM as indicated) for 5 days. Media were collected to assess cell death via LDH release. The low concentrations of Aβ used for these uptake studies did not show any evidence of toxicity.

In order to examine the fate of vesicular TMR-A$\beta_{1\text{-}42}$, after washout, SHSY5Y cells were loaded with TMR-Aβ for 24 hours and then washed with medium (TMR-Aβ was removed from the medium). Within 24-48 hours after wash-out, few fluorescent vesicles remained (Figure 5.3D), suggesting rapid clearance/removal of TMR-Aβ after uptake and trafficking to late endosomes / lysosomes.

5.3.2 Concentration of A$\beta_{1\text{-}42}$ in vesicles

SHSY5Y cells were grown in the presence of 25 nM TMR-A$\beta_{1\text{-}42}$ for 24 hours and confocal microscopy was used to estimate the concentration of TMR-A$\beta_{1\text{-}42}$ in vesicles. A pixel-by-pixel analysis of fluorescence intensity was performed on images captured. Figure 5.5 shows fluorescence intensity on the z-axis (in red) plotted above a high-power two-dimensional image of a group of SHSY5Y cells. The fluorescence within vesicles is considerably higher than extracellular regions of the image.
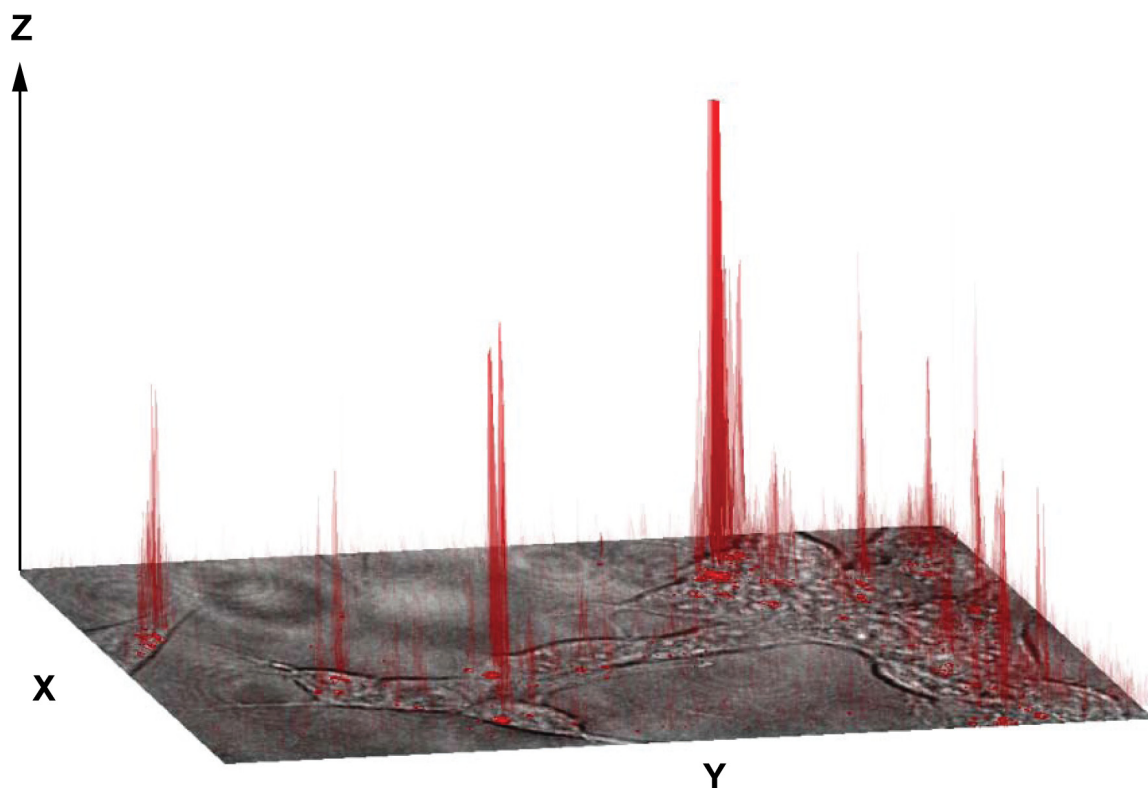
Figure 5.5 Concentrated TMR-Aβ$_{1-42}$ in intracellular vesicles: SHSY5Y cells were grown in the presence of 25 nM TMR-Aβ$_{1-42}$ for 24 hours and imaged using confocal microscopy.   The phase image of a group of cells is plotted in the x-y plane with the fluorescence intensity plotted on the z-axis and projected onto the image.   The fluorescence originating from intracellular vesicles is up to two orders of magnitude greater than fluorescence in the extracellular medium (see Figure 5.6), suggesting high effective concentrations of Aβ in the vesicles.

To obtain a quantitative estimate of Aβ concentration inside the vesicles, pixel intensity was calibrated to known concentrations of free TMR. By comparing pixels from vesicles to the calibration curves, it was possible to estimate the concentration of labeled-Aβ inside the vesicles (Figures 5.6A and 5.6B).   Twenty-nine vesicles from five different

images taken at random positions within the culture dish were analyzed. All vesicles were between 1-3 microns in diameter, a size consistent with both late-endosomes and lysosomes. The concentration of TMR-Aβ_{1-42} in most vesicles was estimated to be between 1 and 2.5 µM; in two of the vesicles the concentration of TMR-Aβ_{1-42} was estimated to be greater than 2.5 µM, although this is also the upper limit of detection with the instrumentation used. These data represent a concentration increase of approximately 100-fold from that originally added to the culture medium. As will be discussed, this is almost certainly an underestimate of the actual concentration inside the vesicles.
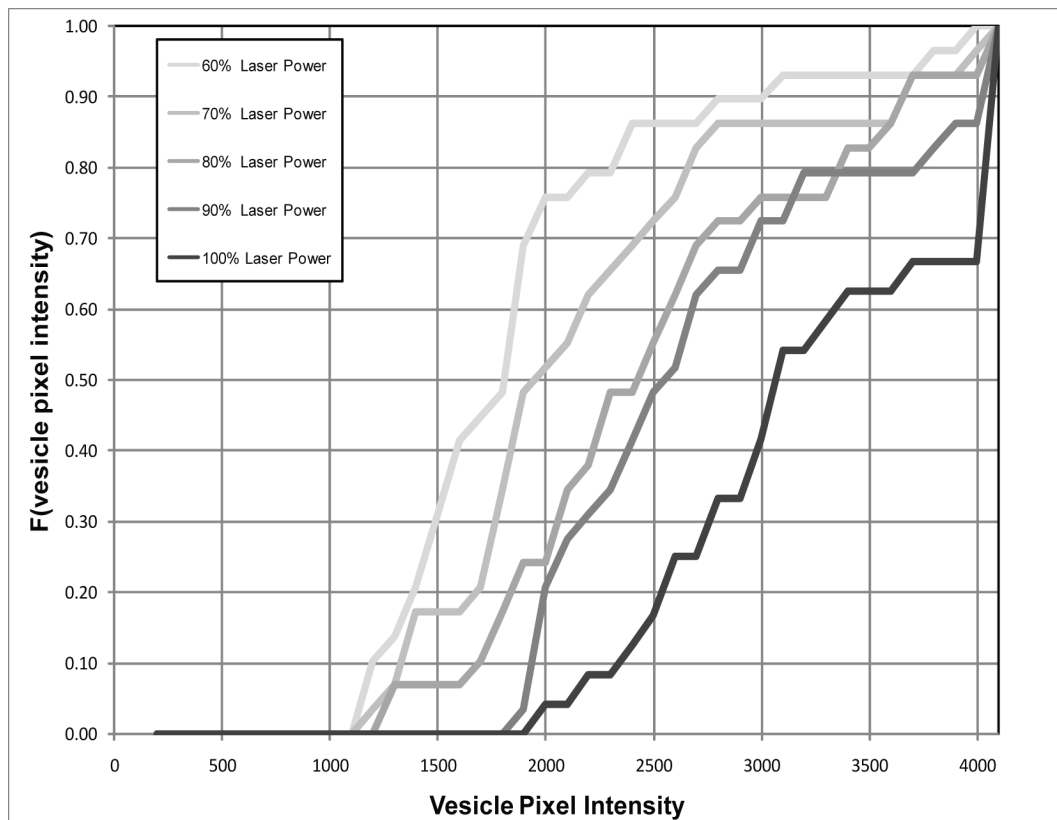
Figure 5.6 A) Pixel intensities were recorded at five different laser excitation powers. Histograms of pixel intensities were converted to cumulative distribution functions

(CDFs). The ordinate shows the value of the CDF, while the abscissa shows pixel intensities. Five curves are shown, one for each of the laser excitation powers used. For specific value of the pixel intensity, the CDF value on the ordinate quantifies the probability of realizing that pixel intensity. The CDF was created from the pixel brightness values observed for 29 vesicles.
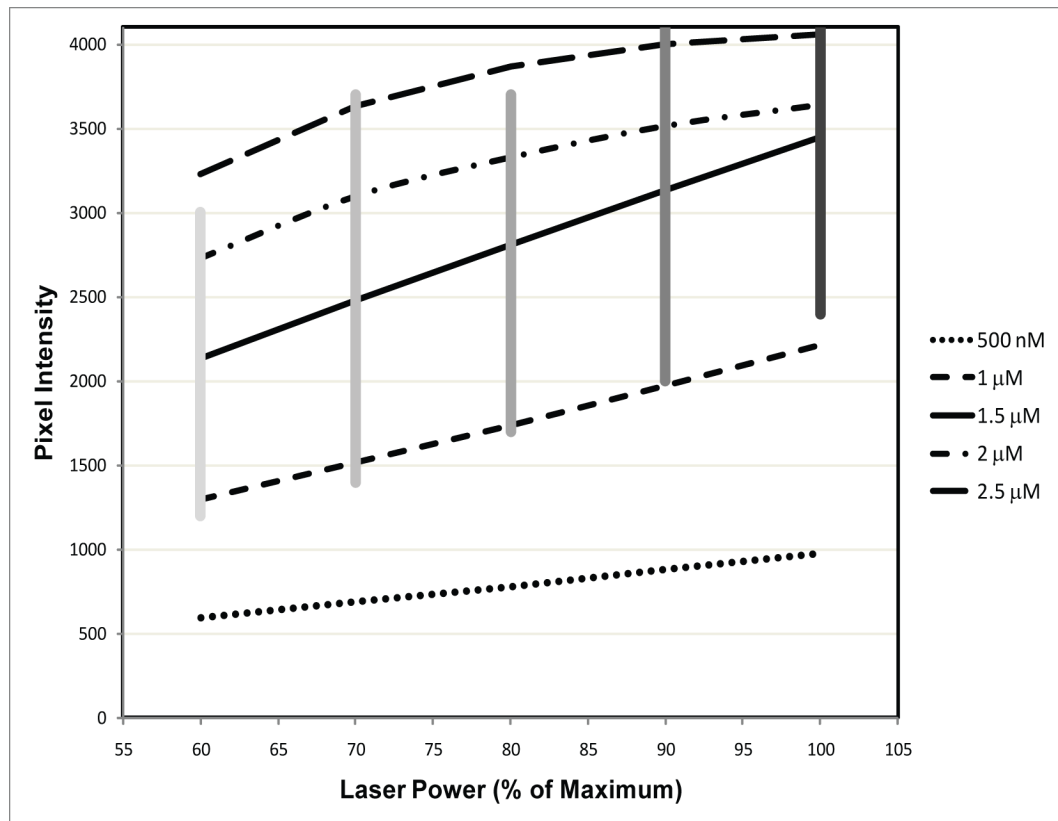
Figure 5.6 B) Vesicle-observed pixel intensities compared to average pixel intensities from dye solutions of known concentration. Vertical lines represent a range from 5% to 95% of the observed pixel intensities from the 29 vesicles analyzed as determined from the cumulative distribution functions in Figure 5.6 A). These are plotted against calibration curves showing the mean observed pixel intensities as a function of laser

power for solutions of known concentrations of free TMR. As can be seen, the observed pixel intensities from the vesicles suggest that the intravesicular concentration is in the micromolar range.

5.3.3 Vesicular $A\beta_{1-42}$ forms high molecular weight aggregates

Given the ~100-fold difference in $A\beta_{1-42}$ concentrations between the extracellular fluid and intracellular vesicles, SHSY5Y cells loaded with $A\beta_{1-42}$ were examined for intracellular aggregates. SHSY5Y cells, grown in the presence of varying concentrations of unlabeled $A\beta_{1-42}$ (0-1000 nM) for 5 days or in the presence of 1 μM $A\beta_{1-42}$ for varying periods of time (0-7 days), were washed, sonicated, and analyzed using agarose gel electrophoresis and immunoblotting with anti-Aβ antibodies (6E10 and 3D6, both N-terminal specific monoclonal antibodies). Tris-tricine gels showed high molecular weight aggregates of Aβ (>200 kDa) in homogenates of cells grown in 1 μM $A\beta_{1-42}$ for 5 days (Figure 5.7A). Moreover, high molecular weight aggregates (>800 kDa) were detected in cell homogenates in a concentration- (Figure 5.7B) and time-dependent (Figure 5.7C) manner, indicating that the intracellular formation of high molecular weight Aβ aggregates occurred only when cells were incubated with high extracellular concentrations of soluble Aβ. The high-molecular weight aggregates were recognized by two different anti-Aβ antibodies (6E10, and 3D6, Figure 5.8). Culture medium from these cells (containing 1 μM $A\beta_{1-42}$) did not show evidence of aggregation in Tris-tricine (Figure 5.7A) or agarose gels (Figure 5.8D), consistent with published results, which have shown that Aβ is less prone to aggregation in serum-containing medium (*36*).

Lower concentrations of extracellular $A\beta_{1-42}$ (0-500 nM) did not result in intracellular aggregation during the 5 day incubation period (Figure 5.7B). Intracellular aggregation appeared to increase with increasing incubation times appearing as early as 2 days (Figure 5.7D). Interestingly, equivalent concentrations of extracellular human $A\beta_{1-40}$ or rat $A\beta_{1-42}$ did not elicit intracellular aggregation (Figure 5.7D).
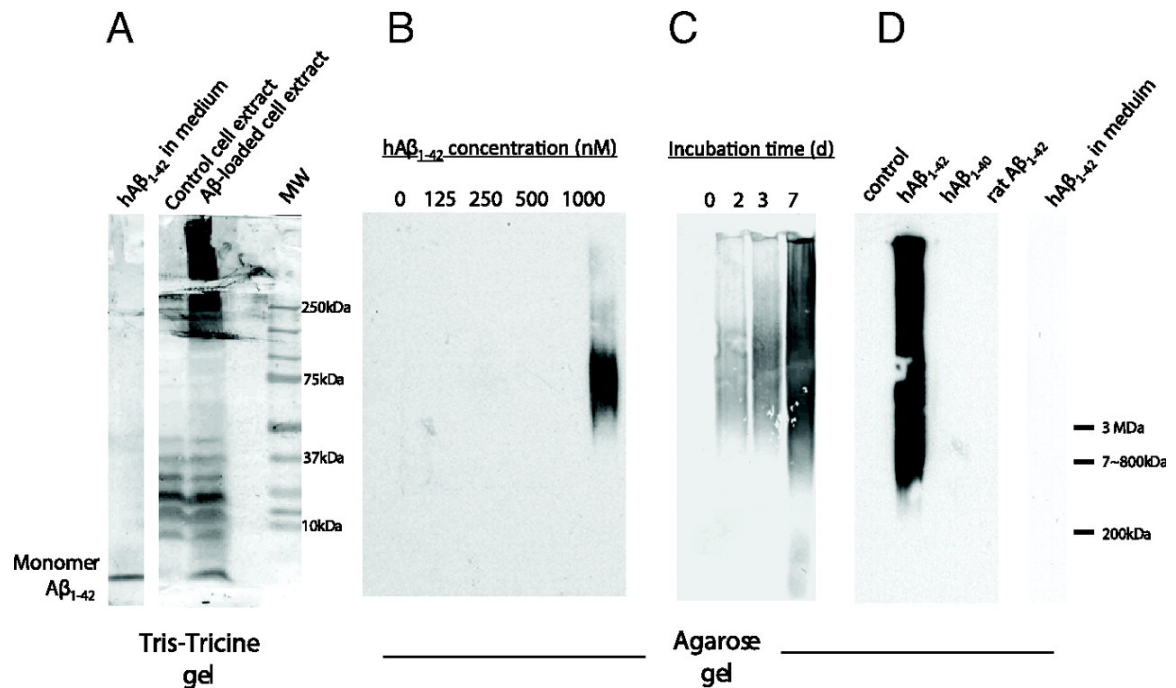
Figure 5.7 Aggregation of intracellular $A\beta_{1-42}$ into high molecular weight forms: A) SHSY5Y cells were grown in the presence or absence of $A\beta_{1-42}$ (1 μM) for 5 days, then homogenized and run on a Tris-Tricine gel and blotted with an anti-$A\beta$ antibody (6E10). Culture medium incubated with $A\beta_{1-42}$ (1 μM) for 5 days show the presence of only monomers. Extracts from cells grown in the presence of $A\beta_{1-42}$ show high molecular weight aggregates, while cells grown without $A\beta$ do not (note the non-specific bands in the 10-40 kDa range from the cell extracts). (B-D) SHSY5Y cells were grown in the

presence of Aβ$_{1-42}$ (0-1000 nM as indicated) for 5 days B), for varying times (0-7 days, 1 μM Aβ) C), or in the presence of human Aβ$_{1-42}$, Aβ$_{1-40}$, or rat Aβ$_{1-42}$ for 5 days (as indicated in D). Cell homogenates were then run on an agarose gel and blotted with an anti-Aβ antibody (6E10). Intracellular Aβ was found to aggregate in a concentration- and time-dependent manner. Human Aβ$_{1-42}$ forms high molecular weight (HMW) intracellular aggregates, but human Aβ$_{1-40}$ and rat Aβ$_{1-42}$ do not. Human Aβ$_{1-42}$ incubated in culture medium alone for 5 days did not form HMW aggregates (right lane).
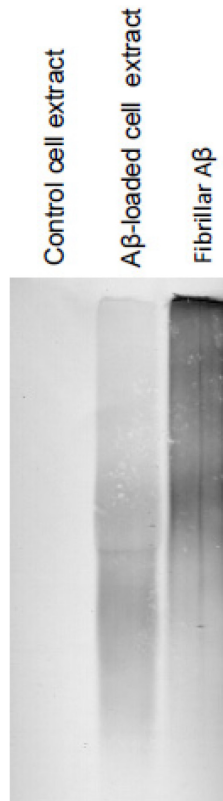
Figure 5.8 SHSY5Y cells were cultured in the presence or absence of 1 μM Aβ$_{1-42}$ for 5 days. Cells were washed, homogenized, and sonicated in 2% SDS and applied to an agarose gel. One lane was loaded with fibrillar Aβ as a positive control. The gel was

transferred to a PDVF membrane and immunoblotted with the 3D6 anti-Aβ antibody (directed at the N-terminus). The immunoblot is identical to that seen using the 6E10 antibody (see Figure 5.7B).

5.3.4 Intracellular Aβ$_{1-42}$ aggregates can seed amyloid fibril formation

To determine if high molecular weight aggregates of Aβ that formed in intracellular vesicles can seed the formation of amyloid fibrils, control cells, cells loaded with unlabeled Aβ$_{1-42}$, and cells incubated with scrambled Aβ$_{1-42}$ were sonicated with SDS, and homogenates were incubated with 100 nM TMR-Aβ$_{1-42}$ at 37°C for 48 hours. This concentration of Aβ is too low for spontaneous formation of fibrils over any time frame in both the intracellular space and the bulk. Aβ-loaded cell extracts induced the appearance of fluorescent precipitates, visible by microscopy (Figure 5.8B), while extracts from cells grown in the absence of Aβ or in the presence of scrambled-Aβ did not (Figure 5.8A). Moreover, these same precipitates stained with Thioflavin-S, suggesting the formation of fibrils (Figure 5.8D). Again, extracts from cells grown in the absence of Aβ or in the presence of scrambled-Aβ failed to develop Thioflavin-S staining (Figure 5.8C).
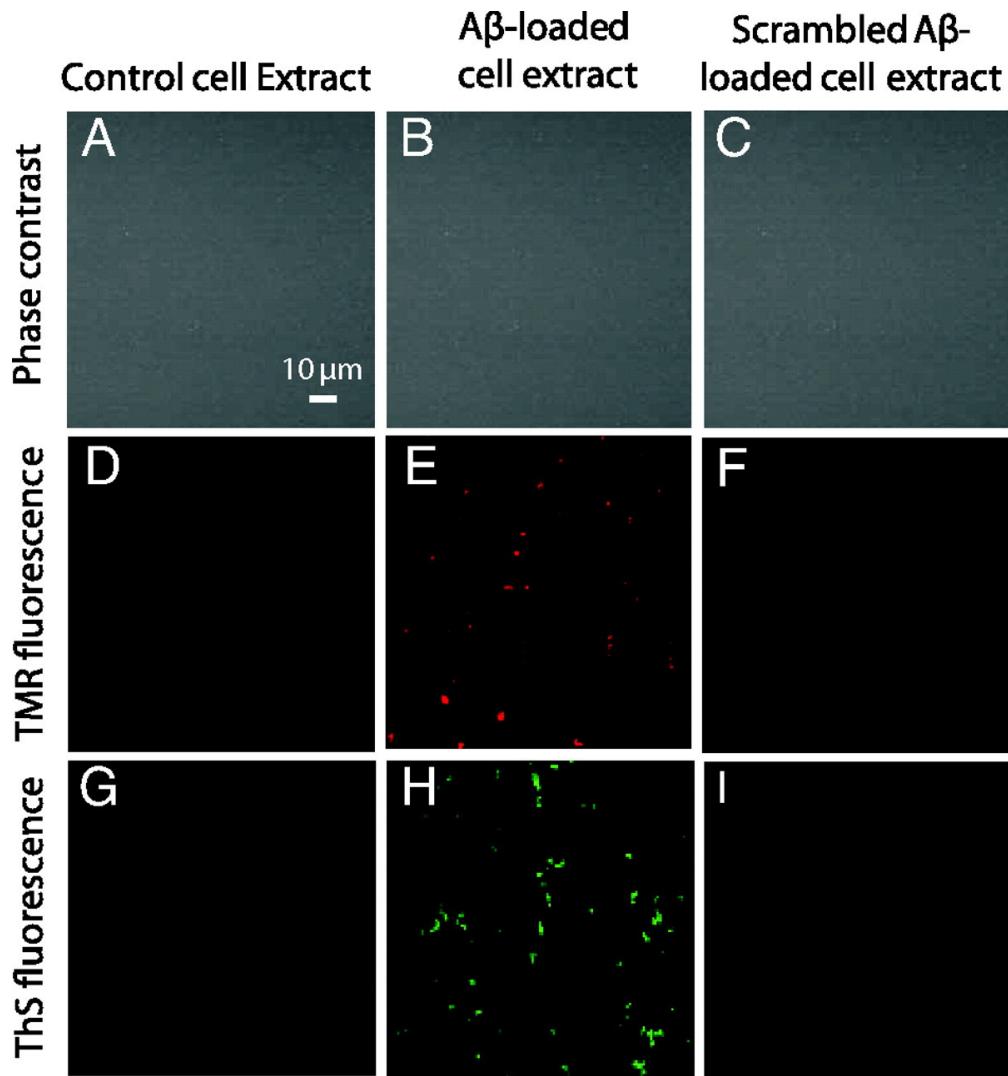
Figure 5.8 Cell extracts from Aβ$_{1-42}$-loaded cells seed the formation of amyloid fibrils. SHSY5Y cells incubated with or without 1 μM unlabeled Aβ$_{1-42}$ or scrambled Aβ$_{1-42}$ for 5 days were homogenized and then incubated with 100 nM TMR-Aβ$_{1-42}$ for 48 hours. Phase contrast images of the cell extracts show the absence of cells (A-C). Extracts from control cells (grown in the absence of Aβ) did not show TMR fluorescence D) or Thioflavin-S staining G); Aβ-loaded cell extracts developed TMR precipitates E) which stained for Thioflavin-S H); while scrambled-Aβ-loaded cell extracts showed neither (F

and I). These results suggest that intracellular Aβ aggregates can seed the formation of Thioflavin-positive aggregates.

5.4 Discussion

A major conundrum in understanding spontaneous appearance of plaque in the brains of Alzheimer's Disease patients is the gap between known extracellular Aβ concentrations *in vivo* (in the low nM range) and the concentration required for spontaneous aggregation *in vitro* (which is in the μM range). This concentration gap spans 3-4 orders of magnitude. In this study, we find that SHSY5Y neuroblastoma cells and cortical neurons are capable of taking up and concentrating extracellular Aβ even at low, physiologically relevant concentrations. Starting with an extracellular concentration of 25 nM, we estimated the concentrations inside some intracellular vesicles to be at least 2.5 μM. However, it is highly likely that this is a conservative estimate for several reasons. First, it is known that adjacent tetramethylrhodamine molecules can self-quench. If there is a high local concentration of fluorophores, the fluorescence could be significantly less than one might expect based on the quantum yield of an individual fluorophore. Second, because fluorescent intensities were quantified in a single 2D plane, there is no guarantee that a given vesicle was entirely contained within the confocal beam volume. Moreover, the fluorescence intensity is maximal at the center of the beam focal volume and decays as e^{-r^2}, where r is the radial distance from the point of maximal intensity, so any offset will lead to significant decreases in observed

fluorescence relative to actual fluorescence. Third, our concentration calibration was performed on a homogeneous solution of fluorophores. This is an ideal case where the assumption of uniform filling of the beam volume is valid. In our experiment, there may be a large number of fluorophores, but they are contained within a volume smaller than that defined by a single pixel. This could be especially pronounced in an aggregate. Therefore, the assumption of fluorophores uniformly filling the beam volume actually leads to an underestimation of the actual concentration. Finally, it is likely that the concentration of $A\beta$ added to the culture medium decreases with time due to the release of $A\beta$-degrading proteases into the culture medium by neural cells (*37*). In light of the preceding discussion, it is reasonable to assume that the local concentrations of $A\beta$ in the intracellular vesicles may, in fact, be an order of magnitude higher than what we have estimated. Therefore, we propose that neural cells may be capable of taking up $A\beta$ and increasing its effective concentration by several orders of magnitude.

Most estimates for the concentration of $A\beta$ in human cerebrospinal fluid (CSF) have been in the low nM range (0.1-1 nM) (*3-7*). Though CSF concentrations of proteins do not necessarily reflect interstitial fluid (ISF) concentrations (the more relevant quantity), a recent study examining $A\beta$ levels in human head trauma patients using microdialysis suggested that the concentration of $A\beta$ is similar in the two pools (*38*). In bulk solutions at nM concentrations, $A\beta$ should exist in soluble form, either as a monomer or low molecular weight oligomer (*39*).

There is growing evidence that Aβ accumulates in vesicles in neurons prior to the development of amyloid plaques. In a variety of AD mouse models, vesicular Aβ immunostaining has been observed in cortical and hippocampal neurons prior to the deposition of amyloid plaques (*40, 41*). Such vesicular Aβ immunostaining is absent in wildtype controls. Similar vesicular Aβ immunostaining has been reported in human AD brains (*42, 43*), in AD-vulnerable brain regions from non-demented individuals, and in brains of young Down's Syndrome cases prior to the deposition of plaques (*42, 44*). Though it has been speculated that this vesicular Aβ represents new Aβ synthesis from APP processing, an equally likely possibility is that this pool represents extracellular Aβ taken up by cells. Our studies focus on cellular uptake of Aβ and do not address cell-autonomous APP processing and Aβ synthesis.

The specific uptake of $Aβ_{1-40}$ and $Aβ_{1-42}$, but not scrambled-Aβ suggests that the uptake mechanism in SHSY5Y cells is receptor-mediated rather than via a non-specific mechanism like bulk uptake (fluorescein uptake was not observed). Although several Aβ-binding receptors have been reported, the low density lipoprotein receptor-like protein-1 (LRP1) is likely a major neuronal Aβ receptor (*45*). LRP1 can bind Aβ either directly (*46*) or via Aβ chaperones such as apoE (*47*). A role for LRP1 in neuronal Aβ uptake is supported by a study demonstrating that TGFβ2-mediated intraneuronal accumulation of brain injected Aβ is markedly inhibited by LRP1 antagonist RAP (*48*). Interestingly, HEK293 cells, which poorly internalized Aβ in this study, express very low levels of LRP1 (data not shown). The cellular uptake mechanisms that lead to intraneuronal Aβ accumulation require further investigation.

196

Why would Aβ be taken up and trafficked to endosomes/lysosomes? The wash-out studies demonstrate that vesicular Aβ rapidly disappears after Aβ is washed out of the culture medium, suggesting an efficient clearance mechanism. However, when continuously exposed to extracellular Aβ, vesicular Aβ accumulates. As vesicular Aβ concentrations increase, it is possible that the degradative capacity of the endosomal/lysosomal pathway may be overwhelmed, leading to the accumulation of Aβ. Such accumulation may result in aggregation, with Aβ aggregates being degraded less efficiently than monomer or small oligomers. Indeed, fibrillar Aβ is much more resistant to proteolysis than monomeric Aβ (*49*). Thus, one might speculate that disruption of this intracellular degradative pathway might lead to enhanced aggregate formation, and, eventually, more plaque deposition. In fact, gene deletion of a major lysosomal protease, cathepsin B, resulted in a dramatic increase in plaque pathogenesis (*50*). Similarly, gene deletion of an endogenous inhibitor of cathepsin, cystatin C, also reduced amyloid plaque load (*51*), and this effect was eliminated in cathepsin B knockout mice (43). These studies add support to the idea that lysosomal proteolytic activity influences amyloid plaque pathogenesis.

Our data suggest that there is a difference between the aggregation behavior of $A\beta_{1-40}$ and $A\beta_{1-42}$ even though both molecules are trafficked into the endosomes / lysosomes. *In vitro* experiments indicate that the saturation concentration for $A\beta_{1-40}$ aggregation is higher than that of $A\beta_{1-42}$ (*10*). This leads to the interesting possibility that aggregation protects $A\beta_{1-42}$ molecules from degradation in endosomes / lysosomes,

whereas the competition between aggregation and degradation favors the latter for $A\beta_{1-40}$. If $A\beta_{1-42}$ molecules form aggregates that are less efficiently degraded in lysosomally-derived vesicles, then these aggregates could accumulate within cells. It is conceivable that these aggregates could then be released into the extracellular space either through active exocytotic mechanisms or through stress-induced cell death.

One of the major questions remaining is: how are plaques formed in the extracellular space? Assuming the precursors to amyloid plaques are formed inside neuronal cells where we suspect the effective $A\beta$ concentration is high enough to support spontaneous aggregation, why do these aggregates not dissolve once they are released into the extracellular space where the $A\beta$ concentrations are low? Moreover, if aggregates of $A\beta$ are released into the extracellular space, they will only act as seeds for aggregation if placed into a supersaturated solution as discussed in Chapter 3. We have already discussed that the concentrations of $A\beta$ found in the extracellular space are sub-saturation (*3-7*) based on *in vitro* estimates (*1, 2*). It is probable that any $A\beta$ aggregates released into the extracellular space will dissolve unless trapped in an aggregated form. Indeed, the senile plaques do not contain only $A\beta$ aggregates, but are instead a smorgasbord of no less than 27 different components. (*52*) The composition of these plaques coupled with what we know about the thermodynamics of the aggregation suggests that they are unlikely to build up over time through accumulation of $A\beta$, but rather, are likely an immune response to cell death which releases $A\beta$ aggregates. This is certainly not a novel hypothesis. (*52-55*)

Future studies on the role of intracellular trafficking and processing of Aβ will focus on the role of the lysosomal environment in facilitating Aβ aggregation and the fate of these aggregates, if they exist. In particular, several questions about the role of the lysosomes are: what is the role of pH; what is the role of finite copy numbers of Aβ; what is the role of the lysosomal membrane surface; what is the role of confinement and macromolecular crowding? Finally, we would like to know: does Aβ actually form aggregates in the cells; if so, are these aggregates exocytosed, degraded, or partitioned into aggresomes (*56, 57*)?

5.5 References

1. Lomakin, A., Teplow, D. B., Kirschner, D. A., and Benedek, G. B. (1997) Kinetic theory of fibrillogenesis of amyloid beta-protein, *Proceedings of the National Academy of Sciences of the United States of America 94*, 7942-7947.
2. Harper, J. D., Wong, S. S., Lieber, C. M., and Lansbury, P. T., Jr. (1999) Assembly of A beta amyloid protofibrils: an in vitro model for a possible early event in Alzheimer's disease, *Biochemistry 38*, 8972-8980.
3. Seubert, P., Vigo-Pelfrey, C., Esch, F., Lee, M., Dovey, H., Davis, D., Sinha, S., Schlossmacher, M., Whaley, J., Swindlehurst, C., and et al. (1992) Isolation and quantification of soluble Alzheimer's beta-peptide from biological fluids, *Nature 359*, 325-327.
4. Galasko, D., Chang, L., Motter, R., Clark, C. M., Kaye, J., Knopman, D., Thomas, R., Kholodenko, D., Schenk, D., Lieberburg, I., Miller, B., Green, R., Basherad, R., Kertiles, L., Boss, M. A., and Seubert, P. (1998) High cerebrospinal fluid tau and low amyloid beta42 levels in the clinical diagnosis of Alzheimer disease and relation to apolipoprotein E genotype, *Archives of neurology 55*, 937-945.
5. Andreasen, N., Hesse, C., Davidsson, P., Minthon, L., Wallin, A., Winblad, B., Vanderstichele, H., Vanmechelen, E., and Blennow, K. (1999) Cerebrospinal fluid beta-amyloid(1-42) in Alzheimer disease: differences between early- and late-onset Alzheimer disease and stability during the course of disease, *Archives of neurology 56*, 673-680.

6.	Strozyk, D., Blennow, K., White, L. R., and Launer, L. J. (2003) CSF Abeta 42 levels correlate with amyloid-neuropathology in a population-based autopsy study, *Neurology 60*, 652-656.

7.	Gustafson, D. R., Skoog, I., Rosengren, L., Zetterberg, H., and Blennow, K. (2007) Cerebrospinal fluid beta-amyloid 1-42 concentration may predict cognitive decline in older women, *Journal of neurology, neurosurgery, and psychiatry 78*, 461-464.

8.	Glenner, G. G., and Wong, C. W. (1984) Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein, *Biochemical and biophysical research communications 120*, 885-890.

9.	Halverson, K., Fraser, P. E., Kirschner, D. A., and Lansbury, P. T., Jr. (1990) Molecular determinants of amyloid deposition in Alzheimer's disease: conformational studies of synthetic beta-protein fragments, *Biochemistry 29*, 2639-2644.

10.	Snyder, S. W., Ladror, U. S., Wade, W. S., Wang, G. T., Barrett, L. W., Matayoshi, E. D., Huffaker, H. J., Krafft, G. A., and Holzman, T. F. (1994) Amyloid-beta aggregation: selective inhibition of aggregation in mixtures of amyloid with different chain lengths, *Biophysical journal 67*, 1216-1228.

11.	Sisodia, S. S. (1992) Secretion of the beta-amyloid precursor protein, *Annals of the New York Academy of Sciences 674*, 53-57.

12.	Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J., and Selkoe, D. J. (2002) Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo, *Nature 416*, 535-539.

13.	Cirrito, J. R., May, P. C., O'Dell, M. A., Taylor, J. W., Parsadanian, M., Cramer, J. W., Audia, J. E., Nissen, J. S., Bales, K. R., Paul, S. M., DeMattos, R. B., and Holtzman, D. M. (2003) In vivo assessment of brain interstitial fluid with microdialysis reveals plaque-associated changes in amyloid-beta metabolism and half-life, *J Neurosci 23*, 8844-8853.

14.	Cirrito, J. R., Yamada, K. A., Finn, M. B., Sloviter, R. S., Bales, K. R., May, P. C., Schoepp, D. D., Paul, S. M., Mennerick, S., and Holtzman, D. M. (2005) Synaptic activity regulates interstitial fluid amyloid-beta levels in vivo, *Neuron 48*, 913-922.

15.	Gorbenko, G. P., and Kinnunen, P. K. (2006) The role of lipid-protein interactions in amyloid-type protein fibril formation, *Chemistry and physics of lipids 141*, 72-82.

16.	Matsuzaki, K. (2007) Physicochemical interactions of amyloid beta-peptide with lipid bilayers, *Biochimica et biophysica acta 1768*, 1935-1942.

17.	Aisenbrey, C., Borowik, T., Bystrom, R., Bokvist, M., Lindstrom, F., Misiak, H., Sani, M. A., and Grobner, G. (2008) How is protein aggregation in amyloidogenic diseases modulated by biological membranes?, *Eur Biophys J 37*, 247-255.

18.	Munishkina, L. A., Cooper, E. M., Uversky, V. N., and Fink, A. L. (2004) The effect of macromolecular crowding on protein aggregation and amyloid fibril formation, *J Mol Recognit 17*, 456-464.

19.   Eriksson, S., Janciauskiene, S., and Lannfelt, L. (1995) Alpha 1-antichymotrypsin regulates Alzheimer beta-amyloid peptide fibril formation, *Proceedings of the National Academy of Sciences of the United States of America 92*, 2313-2317.

20.   Cotman, S. L., Halfter, W., and Cole, G. J. (2000) Agrin binds to beta-amyloid (Abeta), accelerates abeta fibril formation, and is localized to Abeta deposits in Alzheimer's disease brain, *Molecular and cellular neurosciences 15*, 183-198.

21.   Castillo, G. M., Ngo, C., Cummings, J., Wight, T. N., and Snow, A. D. (1997) Perlecan binds to the beta-amyloid proteins (A beta) of Alzheimer's disease, accelerates A beta fibril formation, and maintains A beta fibril stability, *Journal of neurochemistry 69*, 2452-2465.

22.   Snow, A. D., Sekiguchi, R., Nochlin, D., Fraser, P., Kimata, K., Mizutani, A., Arai, M., Schreier, W. A., and Morgan, D. G. (1994) An important role of heparan sulfate proteoglycan (Perlecan) in a model system for the deposition and persistence of fibrillar A beta-amyloid in rat brain, *Neuron 12*, 219-234.

23.   Smith, D. G., Cappai, R., and Barnham, K. J. (2007) The redox chemistry of the Alzheimer's disease amyloid beta peptide, *Biochimica et biophysica acta 1768*, 1976-1990.

24.   Usui, K., Hulleman, J. D., Paulsson, J. F., Siegel, S. J., Powers, E. T., and Kelly, J. W. (2009) Site-specific modification of Alzheimer's peptides by cholesterol oxidation products enhances aggregation energetics and neurotoxicity, *Proceedings of the National Academy of Sciences of the United States of America 106*, 18563-18568.

25.   Gunn, A. P., Masters, C. L., and Cherny, R. A. (2010) Pyroglutamate-Abeta: Role in the natural history of Alzheimer's disease, *Int J Biochem Cell Biol 42*, 1915-1918.

26.   Hortschansky, P., Schroeckh, V., Christopeit, T., Zandomeneghi, G., and Fandrich, M. (2005) The aggregation kinetics of Alzheimer's beta-amyloid peptide is controlled by stochastic nucleation, *Protein Sci 14*, 1753-1759.

27.   Mantyh, P. W., Ghilardi, J. R., Rogers, S., DeMaster, E., Allen, C. J., Stimson, E. R., and Maggio, J. E. (1993) Aluminum, Iron, and Zinc Ions Promote Aggregation of Physiological Concentrations of β-Amyloid Peptide, *Journal of neurochemistry 61*, 1171-1174.

28.   Dolev, I., and Michaelson, D. M. (2006) The nucleation growth and reversibility of Amyloid-beta deposition in vivo, *J Alzheimers Dis 10*, 291-301.

29.   Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A., and Teplow, D. B. (1996) On the nucleation and growth of amyloid beta-protein fibrils: detection of nuclei and quantitation of rate constants, *Proceedings of the National Academy of Sciences of the United States of America 93*, 1125-1129.

30.   Harper, J. D., and Lansbury, P. T., Jr. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins, *Annual review of biochemistry 66*, 385-407.

31.   Mayer, R. J., Landon, M., Lowe, J., Tipler, C., Arnold, J., and Laszlo, L. (1994) Is there a common intracellular bioreactor in which amyloid formation is initiated in neurodegenerative diseases?, *Biochem Soc Trans 22*, 151-155.

32. Knauer, M. F., Soreghan, B., Burdick, D., Kosmoski, J., and Glabe, C. G. (1992) Intracellular accumulation and resistance to degradation of the Alzheimer amyloid A4/beta protein, *Proceedings of the National Academy of Sciences of the United States of America 89*, 7437-7441.

33. Mayer, R. J., Landon, M., Laszlo, L., Lowe, J., and Lennox, G. (1992) Protein processing in lysosomes: the new therapeutic target in neurodegenerative disease, *The Lancet 340*, 156-159.

34. Diehl, H., and Markuszewski, R. (1989) Studies on fluorescein-VII The fluorescence of fluorescein as a function of pH, *Talanta 36*, 416-418.

35. Lakowicz, J. R. (2006) *Principles of fluorescence spectroscopy*, 3rd ed., Springer, New York.

36. Biere, A. L., Ostaszewski, B., Stimson, E. R., Hyman, B. T., Maggio, J. E., and Selkoe, D. J. (1996) Amyloid beta-peptide is transported on lipoproteins and albumin in human plasma, *The Journal of biological chemistry 271*, 32916-32922.

37. Yin, K. J., Cirrito, J. R., Yan, P., Hu, X., Xiao, Q., Pan, X., Bateman, R., Song, H., Hsu, F. F., Turk, J., Xu, J., Hsu, C. Y., Mills, J. C., Holtzman, D. M., and Lee, J. M. (2006) Matrix metalloproteinases expressed by astrocytes mediate extracellular amyloid-beta peptide catabolism, *J Neurosci 26*, 10939-10948.

38. Brody, D. L., Magnoni, S., Schwetye, K. E., Spinner, M. L., Esparza, T. J., Stocchetti, N., Zipfel, G. J., and Holtzman, D. M. (2008) Amyloid-beta dynamics correlate with neurological status in the injured human brain, *Science (New York, N.Y 321*, 1221-1224.

39. Sengupta, P., Garai, K., Sahoo, B., Shi, Y., Callaway, D. J., and Maiti, S. (2003) The amyloid beta peptide (Abeta(1-40)) is thermodynamically soluble at physiological concentrations, *Biochemistry 42*, 10506-10513.

40. Wirths, O., Multhaup, G., Czech, C., Blanchard, V., Moussaoui, S., Tremp, G., Pradier, L., Beyreuther, K., and Bayer, T. A. (2001) Intraneuronal Abeta accumulation precedes plaque formation in beta-amyloid precursor protein and presenilin-1 double-transgenic mice, *Neuroscience letters 306*, 116-120.

41. Oddo, S., Caccamo, A., Shepherd, J. D., Murphy, M. P., Golde, T. E., Kayed, R., Metherate, R., Mattson, M. P., Akbari, Y., and LaFerla, F. M. (2003) Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction, *Neuron 39*, 409-421.

42. Gouras, G. K., Tsai, J., Naslund, J., Vincent, B., Edgar, M., Checler, F., Greenfield, J. P., Haroutunian, V., Buxbaum, J. D., Xu, H., Greengard, P., and Relkin, N. R. (2000) Intraneuronal Abeta42 accumulation in human brain, *The American journal of pathology 156*, 15-20.

43. LaFerla, F. M., Green, K. N., and Oddo, S. (2007) Intracellular amyloid-beta in Alzheimer's disease, *Nature reviews 8*, 499-509.

44. Gyure, K. A., Durham, R., Stewart, W. F., Smialek, J. E., and Troncoso, J. C. (2001) Intraneuronal abeta-amyloid precedes development of amyloid plaques in Down syndrome, *Archives of pathology & laboratory medicine 125*, 489-492.

45. Zerbinatti, C. V., and Bu, G. (2005) LRP and Alzheimer's disease, *Rev Neurosci 16*, 123-135.

46. Deane, R., Wu, Z., Sagare, A., Davis, J., Du Yan, S., Hamm, K., Xu, F., Parisi, M., LaRue, B., Hu, H. W., Spijkers, P., Guo, H., Song, X., Lenting, P. J., Van Nostrand, W. E., and Zlokovic, B. V. (2004) LRP/amyloid beta-peptide interaction mediates differential brain efflux of Abeta isoforms, *Neuron 43*, 333-344.

47. Zerbinatti, C. V., Wahrle, S. E., Kim, H., Cam, J. A., Bales, K., Paul, S. M., Holtzman, D. M., and Bu, G. (2006) Apolipoprotein E and low density lipoprotein receptor-related protein facilitate intraneuronal Abeta42 accumulation in amyloid model mice, *The Journal of biological chemistry 281*, 36180-36186.

48. Harris-White, M. E., Balverde, Z., Lim, G. P., Kim, P., Miller, S. A., Hammer, H., Galasko, D., and Frautschy, S. A. (2004) Role of LRP in TGFbeta2-mediated neuronal uptake of Abeta and effects on memory, *Journal of neuroscience research 77*, 217-228.

49. Soto, C., and Castano, E. M. (1996) The conformation of Alzheimer's beta peptide determines the rate of amyloid formation and its resistance to proteolysis, *The Biochemical journal 314 (Pt 2)*, 701-707.

50. Mueller-Steiner, S., Zhou, Y., Arai, H., Roberson, E. D., Sun, B., Chen, J., Wang, X., Yu, G., Esposito, L., Mucke, L., and Gan, L. (2006) Antiamyloidogenic and neuroprotective functions of cathepsin B: implications for Alzheimer's disease, *Neuron 51*, 703-714.

51. Mi, W., Pawlik, M., Sastre, M., Jung, S. S., Radvinsky, D. S., Klein, A. M., Sommer, J., Schmidt, S. D., Nixon, R. A., Mathews, P. M., and Levy, E. (2007) Cystatin C inhibits amyloid-beta deposition in Alzheimer's disease mouse models, *Nature genetics 39*, 1440-1442.

52. Armstrong, R. A., Lantos, P. L., and Cairns, N. J. (2008) What determines the molecular composition of abnormal protein aggregates in neurodegenerative disease?, *Neuropathology 28*, 351-365.

53. Murphy, M. P., and LeVine, H., 3rd. (2010) Alzheimer's disease and the amyloid-beta peptide, *J Alzheimers Dis 19*, 311-323.

54. Castellani, R. J., Zhu, X., Lee, H.-g., Moreira, P. I., Perry, G., and Smith, M. A. (2007) Neuropathology and treatment of Alzheimer disease: did we lose the forest for the trees?, *Expert Review of Neurotherapeutics 7*, 473-485.

55. Castellani, R. J., Lee, H. G., Zhu, X. W., Perry, G., and Smith, M. A. (2008) Alzheimer disease pathology as a host response, *J Neuropath Exp Neur 67*, 523-531.

56. Garcia-Mata, R., Gao, Y.-S., and Sztul, E. (2002) Hassles with Taking Out the Garbage: Aggravating Aggresomes, *Traffic 3*, 388-396.

57. Kopito, R. R. (2000) Aggresomes, inclusion bodies and protein aggregation, *Trends in Cell Biology 10*, 524-530.

Chapter 6: Summary of Findings and Outlook for Future Work

6.1 Preamble

The work presented in this thesis focused on quantitative studies of phase behavior of aggregation-prone systems such as polyglutamine and implications for aggregation mechanisms. We propose that the work described here provides the foundation for current and future studies aimed at answering important questions regarding polyglutamine aggregation and the connection between this process and the mechanisms of neurodegeneration. In this chapter, we provide a synthesis that summarizes the main findings, a synopsis of the important next steps, and a broader outlook on finding solutions to the vexing problem of understanding the connection between aggregation and neurodegeneration.

6.2 The Main Findings and Implications of this Thesis Work

As noted in Chapter 1, polyglutamine tracts appear to be the only theme that is common to the proteins associated with different polyglutamine expansion diseases. (*1-6*) Proteolytic products of these proteins encompass the polyglutamine expansions and these tracts end up as neuronal intranuclear inclusions. (*7-11*) The insolubility of polar polyamide tracts was confounding until a combination of computer simulations from the Pappu lab (*12-15*) and fluorescence correlation spectroscopy measurements (*16*) summarized in Chapter 2 provided evidence that aqueous milieus are poor solvents for polymers such as polyglutamine. This is not just a qualitative statement. Instead, one can gather quantitative evidence for this as shown in both Chapters 2 and 3.

The phase diagrams and AFM data presented in Chapter 3 suggest that the aggregation of polyglutamine is likely to involve heterogeneities. This is inconsistent with the prevalent view in the polyglutamine literature that polyglutamine aggregation is a homogeneous nucleation process with a critical nucleus size of one molecule. (*17-19*) In Chapter 4, we used classical nucleation theory (*20-25*) to reanalyze the kinetic data on which this claim is based. The results suggest that the nucleation is better characterized as heterogeneous nucleation. This is consistent with both the phase diagrams and AFM images presented in Chapter 3 and work from Vitalis and Pappu (*26*).

6.3 Future Work

6.3.1 Unanswered Thermodynamic Questions

From a biophysical standpoint, the question of why water should be a poor solvent for polar polyamides such as polyglutamine remains inadequately answered. Answering this question will require the use of large-scale computer simulations combined with solvent perturbation experiments that attempt to make perturbations or large-scale changes to solvent quality. Interestingly, we invested significant effort in trying to denature polyglutamine tracts using high concentrations of urea and guanidinium hydrochloride. The results depended on chain length and longer chains ($N >$ ~30) were surprisingly resistant to chemical denaturation. As seen in Chapter 3, the inferred UCST is deep in the superheated regime suggesting that the thermal stability of compact and aggregated polyglutamine is considerably higher than soluble proteins of comparable molecular weight. (*27*) Quantitative approaches aimed at understanding how and why polyamides collapse and are resistant to destabilizing forces are important because they will likely lead us to figuring out ways to disrupt aggregates that are held together via strong interactions between polyamide molecules.

6.3.2 Unresolved Mechanistic Issues

The phase diagrams presented in Chapter 3 represent our initial attempts at generating mechanistic insights from quantitative studies of phase behavior. The large metastablity regions seen in these phase diagrams are suggestive of significant

heterogeneities in the process of aggregation that leads to phase separation. There are five

immediate tasks that lie ahead in terms of building on these phase diagrams:

1. *Need for a theoretical formalism for understanding the coupling between Δ_M*

   and aggregation mechanisms: As we discussed in Chapter 3, the combination

   of starting concentrations and solution conditions (pH, salt concentration,

   buffer conditions, temperature and pressure) for most biophysical experiments

   on polyglutamine aggregation are consistent with a nucleation-dependent

   process because the degree of supersaturation ensures that the systems lie

   between the saturation curve and inferred spinodal. (*28*) Given the degree of

   metastability as inferred by values for Δ_M, it would be difficult to conceive of

   a truly homogeneous nucleation mechanism being the only route to phase

   separation. However, the lack of a theoretical formalism to allow quantitative

   connections between Δ_M and the expected degree of heterogeneity in terms of

   the diverse set of pathways and species to be expected is unsatisfying. It

   appears possible to remedy this deficiency through theoretical advancements

   that include order parameters such as density fluctuations. Additionally,

   measuring aggregation kinetics at different temperatures should provide

   phenomenological inputs for understanding the linkage between Δ_M and the

   degree of homogeneity / heterogeneity in aggregation mechanisms. This

   linkage is both important and relevant because it appears that the large values

   we observe for Δ_M provide a thermodynamic rationalization for the

observation of oligomers in cellular and *in vitro* studies of polyglutamine aggregation. (*29-38*)

2. *Is there a liquid-liquid demixing (39) boundary within the metastability region*? Our AFM data suggest the formation of early clusters that are morphologically distinct from the fibrillar morphologies one observes during the late stages of aggregation. Although considerable effort has been directed at structural characterization of oligomers/clusters (*40-45*), it is reasonable postulate that the organization of chains around each other within these clusters is liquid-like. Indeed, it is also possible that the large metastability interval encompasses a liquid-liquid demixing boundary. This should involve a transition from a solution of well-dispersed globules to a solution comprising of a heterogeneous distribution of oligomers and larger clusters that we have referred to as mesoglobules. A combination of laser light scattering measurements and AFM measurements of the type that were introduced in Chapter 3 should provide a direct test of this hypothesis. The nature of the "nucleated conformational conversion" within oligomers/clusters also needs close scrutiny. Advanced AFM methods appear well suited for interrogating this process in conjunction with atomistic and mescoscopic simulations.

3. *Determining the effects of naturally-occurring, flanking sequences on C_{sat} and Δ_M*: We showed that C_{sat} shifts toward lower values as the number of lysine

208

residues is reduced for a given polyglutamine length. This points to stabilizing effects of flanking lysine residues in synthetic peptides. Of course, proteins associated with disease have distinct flanking sequences and some of these remain attached to polyglutamine tracts after proteolytic processing. (*7-11, 46, 47*) If the flanking sequences have a so-called "gatekeeping" effect, then these sequences should shift the saturation curves toward higher concentrations thereby having a direct effect on the driving force for phase separation. Another alternative is that Δ_M will decrease – for a particular polyglutamine tract length – in the context of specific types of flanking sequences. Indeed such an effect is seen even with the addition of an extra pair of lysine residues at the N-terminus (see Chapter 3). If this were the case, then one would expect a diminution in the degree of heterogeneity in the mechanism of aggregation, leading to well-defined on-pathway oligomeric intermediates that turn over into insoluble fibrillar forms. Reducing the degree of heterogeneity by reducing Δ_M might have an indirect gatekeeping effect *in vivo* because it quite possibly reduces the deleterious interactions with an assortment of species or reduces the degree to which other glutamine-rich proteins are recruited into the aggregation process. If flanking sequences increase the driving force for aggregation / phase separation, then the effect is likely to be manifested as a shift of C_{sat} toward lower values, an increase in Δ_M, or both. Combining the measurements described in Chapter 3 with additional probes and computer simulations will be essential for resolving the

debate that surrounds the role of naturally occurring flanking sequences on polyglutamine aggregation.

4. *Crossing the saturation boundary in vivo:* A grand challenge in the protein aggregation field revolves around issue of a saturation concentration paradox. Most measurements of the *in vivo* concentrations of aggregation-prone molecules suggest that these values are in the picomolar / nanomolar range. (*48-51*) These concentrations are clearly below the saturation concentrations that have been measured to date for systems such as polyglutamine, *in vitro*. (*19, 52*) And yet, insoluble inclusions form *in vivo* and the process of aggregation is connected in unresolved ways to the onset of neurodegeneration. What are the mechanisms by which the saturation concentration threshold is crossed? From a physical standpoint, chemical modifications to aggregation-prone molecules and changes in the solution environment can lower the saturation concentrations and bring them in line with physiologically relevant levels. (*53-55*) Alternatively, concentration of these molecules into small volumes, i.e., confinement or interactions with surfaces (*56*) i.e., adsorption can lead to effective / local concentrations or C_{eff} that are greater than C_{sat} and provide a driving force for phase separation. If aggregation occurs through increases in effective concentrations, the mechanisms are unlikely to be akin to those in bulk volumes. This is because the contributions due to fluctuations in composition / density are likely to be

considerably larger than they would be in a bulk setting. In such a scenario, the leading contributions to the driving forces for phase separation comes from fluctuations and the mechanism could end up being diffusion-limited (spinodal decomposition) as opposed to being limited by specific barriers. Detailed biophysical studies of the effects of confinement and membrane adsorption will be invaluable in terms of discerning the principles of aggregation *in vivo* that appear to be governed by different rules or at least fall under different regimes when compared to the *in vitro* setting.

5. *Techniques to map the proposed-rate determining conformational conversion within oligomers:* Serio and Lindquist first proposed the idea of nucleated conformational conversion within oligomers followed by turnover of oligomers to conformations that serve as templates for addition of monomers. (*57*) Numerical models developed by Vitalis and Pappu (*26*) as well as computer simulations from the Pappu lab (*14, 15, 58*) suggest that beta sheet formation is an emergent property in that it is not intrinsic to individual aggregation-prone polypeptides, but is instead a property that arises from the favorable intermolecular interfaces that are realized in "self-solvated" assemblies such as oligomers and / or mesoglobules. However, to date, this model remains in the realm of a plausible proposal supported by indirect data. The challenge for experiments and computer simulations is to find ways to follow the proposed conformational transitions within oligomeric assemblies.

This is difficult because all of the methods that are deployed to study conformational changes are based on averaging over ensembles of molecules. (*59*) Clearly, there is a need for devising single-molecule based strategies to probe conformational changes in a way that is analogous to electronic circular dichroism that allows the ability to distinguish different secondary structures. Conventional measures such as end-to-end distances are reasonably useful, but need to be complemented by clear discriminators between distinct types of structures.

6.4 Other Contributions

In addition to the main contents of this thesis, I contributed to several Pappu lab projects (*58, 60*) as the only resident experimentalist who provided the group of computational biophysicists the ability to test specific predictions that resulted from computer simulations. These collaborative efforts were mutually beneficial because they allowed the Pappu lab to establish an approach that routinely involves closing the loop with experiments while giving me the opportunity to test compelling predictions and questions from results that emerged from simulation work.

In this sense, I believe that my thesis experience has established a new mode of training of experimentalists because it has engrained in me a deep appreciation for theory and computation while helping push my theory / computation colleagues to challenge

themselves to convert some of their findings into predictions that lead to direct and quantitative assessments by experiment. In addition to being intellectually satisfying, the collaborations have given me the opportunity to be involved in a range of projects that have turned into two co-authored papers that I discuss in the appendix to this thesis and two manuscripts that are currently in preparation.

6.5 References

1. Rubinsztein, D. C., Wyttenbach, A., and Rankin, J. (1999) Intracellular inclusions, pathological markers in diseases caused by expanded polyglutamine tracts?, *J Med Genet 36*, 265-270.
2. Margolis, R. L., McInnis, M. G., Rosenblatt, A., and Ross, C. A. (1999) Trinucleotide repeat expansion and neuropsychiatric disease, *Arch Gen Psychiatry 56*, 1019-1031.
3. Bates, G. P., Mangiarini, L., Wanker, E. E., and Davies, S. W. (1998) Polyglutamine expansion and Huntington's disease, *Biochem Soc Trans 26*, 471-475.
4. Trinka, E., Luthringshausen, G., Ladurner, G., Waigell-Weber, M., and Spiegel, R. (1996) Correlations between triplet repeat expansion and clinical features in Huntington's disease, *Arch Neurol 53*, 714-715.
5. Saudou, F., Devys, D., Trottier, Y., Imbert, G., Stoeckel, M. E., Brice, A., and Mandel, J. L. (1996) Polyglutamine expansions and neurodegenerative diseases, *Cold Spring Harb Symp Quant Biol 61*, 639-647.
6. La Spada, A. R., Paulson, H. L., and Fischbeck, K. H. (1994) Trinucleotide repeat expansion in neurological disease, *Ann Neurol 36*, 814-822.
7. Colomer Gould, V. F., Goti, D., Pearce, D., Gonzalez, G. A., Gao, H., Bermudez de Leon, M., Jenkins, N. A., Copeland, N. G., Ross, C. A., and Brown, D. R. (2007) A Mutant ataxin-3 fragment results from processing at a site N-terminal to amino acid 190 in brain of Machado-Joseph disease-like transgenic mice, *Neurobiology of Disease 27*, 362-369.
8. Hoffner, G., Island, M. L., and Djian, P. (2005) Purification of neuronal inclusions of patients with Huntington's disease reveals a broad range of N-terminal fragments of expanded huntingtin and insoluble polymers, *Journal of Neurochemistry 95*, 125-136.
9. Venkatraman, P., Wetzel, R., Tanaka, M., Nukina, N., and Goldberg, A. L. (2004) Eukaryotic Proteasomes Cannot Digest Polyglutamine Sequences and Release

Them during Degradation of Polyglutamine-Containing Proteins, *Molecular Cell 14*, 95-104.

10. Sun, B., Fan, W., Balciunas, A., Cooper, J. K., Bitan, G., Steavenson, S., Denis, P. E., Young, Y., Adler, B., Daugherty, L., Manoukian, R., Elliott, G., Shen, W., Talvenheimo, J., Teplow, D. B., Haniu, M., Haldankar, R., Wypych, J., Ross, C. A., Citron, M., and Richards, W. G. (2002) Polyglutamine repeat length-dependent proteolysis of huntingtin, *Neurobiol Dis 11*, 111-122.

11. Kobayashi, Y., Miwa, S., Merry, D. E., Kume, A., Mei, L., Doyu, M., and Sobue, G. (1998) Caspase-3 cleaves the expanded androgen receptor protein of spinal and bulbar muscular atrophy in a polyglutamine repeat length-dependent manner, *Biochem Biophys Res Commun 252*, 145-150.

12. Wang, X., Vitalis, A., Wyczalkowski, M. A., and Pappu, R. V. (2006) Characterizing the conformational ensemble of monomeric polyglutamine, *Proteins 63*, 297-311.

13. Vitalis, A., Wang, X., and Pappu, R. V. (2007) Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories, *Biophys J 93*, 1923-1937.

14. Vitalis, A., Wang, X., and Pappu, R. V. (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization, *J Mol Biol 384*, 279-297.

15. Vitalis, A., Lyle, N., and Pappu, R. V. (2009) Thermodynamics of beta-sheet formation in polyglutamine, *Biophys J 97*, 303-311.

16. Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R., and Pappu, R. V. (2006) Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions, *Proc Natl Acad Sci U S A 103*, 16764-16769.

17. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction, *Proc Natl Acad Sci U S A 102*, 15400-15405.

18. Chen, S., Ferrone, F. A., and Wetzel, R. (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation, *Proc Natl Acad Sci U S A 99*, 11884-11889.

19. Kar, K., Jayaraman, M., Sahoo, B., Kodali, R., and Wetzel, R. (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent, *Nat Struct Mol Biol 18*, 328-336.

20. Turnbull, D., and Vonnegut, B. (1952) *Ind. Eng. Chem. 44*, 1292.

21. Turnbull, D. (1950) *J. Chem. Phys. 18*, 198.

22. Sear, R. P. (2007) Nucleation: theory and applications to protein solutions and colloidal suspensions, *Journal of Physics: Condensed Matter 19*, 033101.

23. Oxtoby, D. W. (1998) *Acc. Chem. Res. 31*, 91.

24. Oxtoby, D. W. (1992) *J. Phys.: Condens. Matter 4*, 7627.

25. Debenedetti, P. G. (1996) *Metastable Liquids*.

26. Vitalis, A., and Pappu, R. V. (2011) Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides, *Biophys Chem*.

27. Pace, C. N., Trevino, S., Prabhakaran, E., and Scholtz, J. M. (2004) Protein structure, stability and solubility in water and other solvents, *Philos Trans R Soc Lond B Biol Sci 359*, 1225-1234; discussion 1234-1225.

28. Rubinstein, M., and Colby, R. H. (2003) *Polymer Physics*, Oxford University Press, Oxford and New York.

29. Lee, C. C., Walters, R. H., and Murphy, R. M. (2007) Reconsidering the mechanism of polyglutamine peptide aggregation, *Biochemistry 46*, 12810-12820.

30. Hands, S., and Wyttenbach, A. (2010) Neurotoxic protein oligomerisation associated with polyglutamine diseases, *Acta Neuropathologica 120*, 419-437.

31. Legleiter, J., Mitchell, E., Lotz, G. P., Sapp, E., Ng, C., DiFiglia, M., Thompson, L. M., and Muchowski, P. J. (2010) Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner in vitro and in vivo, *J Biol Chem 285*, 14777-14790.

32. Lotz, G. P., Legleiter, J., Aron, R., Mitchell, E. J., Huang, S. Y., Ng, C., Glabe, C., Thompson, L. M., and Muchowski, P. J. (2010) Hsp70 and Hsp40 functionally interact with soluble mutant huntingtin oligomers in a classic ATP-dependent reaction cycle, *J Biol Chem 285*, 38183-38193.

33. Shirendeb, U., Reddy, A. P., Manczak, M., Calkins, M. J., Mao, P., Tagle, D. A., and Reddy, P. H. (2011) Abnormal mitochondrial dynamics, mitochondrial loss and mutant huntingtin oligomers in Huntington's disease: implications for selective neuronal damage, *Hum Mol Genet 20*, 1438-1455.

34. Lajoie, P., and Snapp, E. L. (2010) Formation and toxicity of soluble polyglutamine oligomers in living cells, *PLoS One 5*, e15245.

35. Takahashi, T., Kikuchi, S., Katada, S., Nagai, Y., Nishizawa, M., and Onodera, O. (2008) Soluble polyglutamine oligomers formed prior to inclusion body formation are cytotoxic, *Hum Mol Genet 17*, 345-356.

36. Takahashi, Y., Okamoto, Y., Popiel, H. A., Fujikake, N., Toda, T., Kinjo, M., and Nagai, Y. (2007) Detection of polyglutamine protein oligomers in cells by fluorescence correlation spectroscopy, *J Biol Chem 282*, 24039-24048.

37. Wacker, J. L., Zareie, M. H., Fong, H., Sarikaya, M., and Muchowski, P. J. (2004) Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer, *Nat Struct Mol Biol 11*, 1215-1222.

38. Wong, S. L., Chan, W. M., and Chan, H. Y. (2008) Sodium dodecyl sulfate-insoluble oligomers are involved in polyglutamine degeneration, *FASEB J 22*, 3348-3357.

39. Padrick, S. B., and Miranker, A. D. (2002) Islet amyloid: phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis, *Biochemistry 41*, 4694-4703.

40. Eslami, H., and Muller-Plathe, F. (2009) Structure and mobility of nanoconfined polyamide-6,6 oligomers: application of a molecular dynamics technique with

constant temperature, surface area, and parallel pressure, *J Phys Chem B 113*, 5568-5581.

41. Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis, *Science 300*, 486-489.

42. Ono, K., Condron, M. M., and Teplow, D. B. (2009) Structure-neurotoxicity relationships of amyloid beta-protein oligomers, *Proc Natl Acad Sci U S A 106*, 14745-14750.

43. Pitt, J., Roth, W., Lacor, P., Smith, A. B., 3rd, Blankenship, M., Velasco, P., De Felice, F., Breslin, P., and Klein, W. L. (2009) Alzheimer's-associated Abeta oligomers show altered structure, immunoreactivity and synaptotoxicity with low doses of oleocanthal, *Toxicol Appl Pharmacol 240*, 189-197.

44. Rohrig, U. F., Laio, A., Tantalo, N., Parrinello, M., and Petronzio, R. (2006) Stability and structure of oligomers of the Alzheimer peptide Abeta16-22: from the dimer to the 32-mer, *Biophys J 91*, 3217-3229.

45. Walsh, P., Yau, J., Simonetti, K., and Sharpe, S. (2009) Morphology and secondary structure of stable beta-oligomers formed by amyloid peptide PrP(106-126), *Biochemistry 48*, 5779-5781.

46. Cooper, J. K., Schilling, G., Peters, M. F., Herring, W. J., Sharp, A. H., Kaminsky, Z., Masone, J., Khan, F. A., Delanoy, M., Borchelt, D. R., Dawson, V. L., Dawson, T. M., and Ross, C. A. (1998) Truncated N-Terminal Fragments of Huntingtin with Expanded Glutamine Repeats form Nuclear and Cytoplasmic Aggregates in Cell Culture, *Human Molecular Genetics 7*, 783-790.

47. Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J. L., Fischbeck, K. H., and Pittman, R. N. (1997) Intranuclear Inclusions of Expanded Polyglutamine Protein in Spinocerebellar Ataxia Type 3, *Neuron 19*, 333-344.

48. Gustafson, D. R., Skoog, I., Rosengren, L., Zetterberg, H., and Blennow, K. (2007) Cerebrospinal fluid beta-amyloid 1-42 concentration may predict cognitive decline in older women, *Journal of neurology, neurosurgery, and psychiatry 78*, 461-464.

49. Cirrito, J. R., Yamada, K. A., Finn, M. B., Sloviter, R. S., Bales, K. R., May, P. C., Schoepp, D. D., Paul, S. M., Mennerick, S., and Holtzman, D. M. (2005) Synaptic activity regulates interstitial fluid amyloid-beta levels in vivo, *Neuron 48*, 913-922.

50. Strozyk, D., Blennow, K., White, L. R., and Launer, L. J. (2003) CSF Abeta 42 levels correlate with amyloid-neuropathology in a population-based autopsy study, *Neurology 60*, 652-656.

51. Cirrito, J. R., May, P. C., O'Dell, M. A., Taylor, J. W., Parsadanian, M., Cramer, J. W., Audia, J. E., Nissen, J. S., Bales, K. R., Paul, S. M., DeMattos, R. B., and Holtzman, D. M. (2003) In vivo assessment of brain interstitial fluid with microdialysis reveals plaque-associated changes in amyloid-beta metabolism and half-life, *J Neurosci 23*, 8844-8853.

52. Chen, S., Berthelier, V., Hamilton, J. B., O'Nuallain, B., and Wetzel, R. (2002) Amyloid-like features of polyglutamine aggregates and their assembly kinetics, *Biochemistry 41*, 7391-7399.

53. Gunn, A. P., Masters, C. L., and Cherny, R. A. (2010) Pyroglutamate-Abeta: Role in the natural history of Alzheimer's disease, *Int J Biochem Cell Biol 42*, 1915-1918.

54. Usui, K., Hulleman, J. D., Paulsson, J. F., Siegel, S. J., Powers, E. T., and Kelly, J. W. (2009) Site-specific modification of Alzheimer's peptides by cholesterol oxidation products enhances aggregation energetics and neurotoxicity, *Proceedings of the National Academy of Sciences of the United States of America 106*, 18563-18568.

55. Hortschansky, P., Schroeckh, V., Christopeit, T., Zandomeneghi, G., and Fandrich, M. (2005) The aggregation kinetics of Alzheimer's beta-amyloid peptide is controlled by stochastic nucleation, *Protein Sci 14*, 1753-1759.

56. Gorbenko, G. P., and Kinnunen, P. K. (2006) The role of lipid-protein interactions in amyloid-type protein fibril formation, *Chemistry and physics of lipids 141*, 72-82.

57. Krishnan, R., and Lindquist, S. L. (2005) Structural insights into a yeast prion illuminate nucleation and strain diversity, *Nature 435*, 765-772.

58. Williamson, T. E., Vitalis, A., Crick, S. L., and Pappu, R. V. (2010) Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin, *J Mol Biol 396*, 1295-1309.

59. Frieden, C. (2007) Protein aggregation processes: In search of the mechanism, *Protein Sci 16*, 2334-2344.

60. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., and Pappu, R. V. (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins, *Proc Natl Acad Sci U S A 107*, 8183-8188.

Appendix: Compendium of Personal Contributions to other Scientific Works

Aside from the work presented in the thesis, I have also contributed to other scientific works in both the Pappu lab and Frieden lab. I will list the references to these publications, and comment on my personal contributions to them:

1. **Garai, K., Crick, S. L., Mustafi, S. M., and Frieden, C. (2009) Expression and purification of amyloid-beta peptides from Escherichia coli,** *Protein Expr Purif 66*, **107-112.**

In this work, we created a fusion protein of amyloid-beta and intestinal fatty-acid binding protein. A cleavable linker was inserted between the fusion so we could recover pure amyloid-beta after cleavage. We hoped this system would allow us to make amyloid-beta recombinantly and in good yield. My contributions to this work were realized in the development of a purification strategy of the amyloid-beta protein after cleavage. Specifically, I developed HPLC protocols to assist in the purification.

2. **Williamson, T. E., Vitalis, A., Crick, S. L., and Pappu, R. V. (2010) Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin,** *J Mol Biol 396*, **1295-1309.**

Tim Williamson from the Pappu Lab simulated polyglutamine molecules with and without the N-terminal sequence (N17) from the huntingtin protein. One of the observations he made was that the % helix of the N17 was modulated by the amount of

glutamine it was attached to. He calculated the N17 alone to have ~33% helix. My

contribution to this work was to test this prediction using circular dichroism experiments.

I titrated both TFE and urea to obtain the spectral baselines for 100 % helix and 100 %

coil, respectively. I then measured the spectrum of the N17 in absence of these titrants.

Because I had the baselines for 100% helix and 100% coil, I could decompose the signal

measured in aqueous buffer into percentages of helix and coil. I calculated this peptide to

contain about 34% helix in aqueous buffer, which was consistent with the simulations.

3. **Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., and Pappu, R. V. (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins, *Proc Natl Acad Sci U S A 107*, 8183-8188.**

Albert Mao from the Pappu Lab performed simulations of protamine-like

molecules varying in net-charge per residue from close to 0 to 1 for polyarginine. He

observed that peptides with a net charge below 0.28 adopted collapsed conformations on

average while peptides with a net charge above 0.28 adopted random-coil conformations.

My contribution to this work was to help design and perform experiments to test this

prediction. I purified synthesized peptides, labeled them with fluorescent dyes, and

performed both fluorescence correlation spectroscopy (FCS) and fluorescence anisotropy

experiments. These experiments recapitulated what Albert had predicted based on his

simulations. Importantly, Albert and I worked together to determine a way in which we

could obtain quantitative agreement with the FCS experiments using HYDROPRO to

calculate an average diffusion coefficient from his simulations. The parameters of

HYDROPRO were adjusted to give the correct diffusion coefficient for

tetramethylrhodamine, which I measured with FCS. This was an important technique

that arose from this work, and one which we will likely use in the future to connect predictions from simulation to experimental measurements.

4. **Mustafi, S. M., Garai, K., Crick, S. L., Baban, B., and Frieden, C. (2010) Substoichiometric inhibition of Abeta(1-40) aggregation by a tandem Abeta(40-1-Gly8-1-40) peptide,** *Biochem Biophys Res Commun 397***, 509-512.**

The motivation for this work was to create a possible seed for amyloid beta aggregation by making a dimer of the amyloid-beta peptide by linking the C-terminus of one molecule to the N-terminus of the next. My contribution to this work was that the construction of a dimeric Abeta in this fashion was my original idea.

Scott L. Crick

Graduate Student in the labs of
Carl Frieden and Rohit Pappu

Washington University in St. Louis
One Brookings Drive
Campus Box 1097
St. Louis, Missouri 63130
cell: (314) 435-8654
scrick@gmail.com

EDUCATION

Washington University in St. Louis
Ph.D. in Biomedical Engineering
August 2011
Thesis: Inferring Aggregation Mechanisms of Molecules Involved in Neurodegeneration through Quantitative Studies of Phase Behavior

Washington University in St. Louis
B.S. in Biomedical Engineering (*Summa Cum Laude*)
May 2005
Focus in Biomolecular Engineering

AWARDS & HONORS

NSF Graduate Research Fellow, *2005-2008*
Bachelor of Science degree (Summa Cum Laude), *2005*
NIH Summer Research Fellow at Washington University, *2003*
Biomedical Engineering Outstanding Junior Award, *2003*
HHMI Summer Research Fellow at Washington University, *2002*
Antoinette Frances Dames Award for Academic Excellence, *2002*
Annual Dean's Award Undergraduate Scholarship, *Fall 2001-Spring 2005*
Washington University Dean's List, *Fall 2001-Spring 2005*
HHMI Prefreshman Summer Scholar at Washington University, *2001*
Henderson County High School Valedictorian, *2001*

PUBLICATIONS

Research Papers:

1. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. Proc Natl Acad Sci U S A. 2006;103(45):16764-9. PMCID: 1629004.

2. Crick SL, Yin FC. Assessing micromechanical properties of cells with atomic force microscopy: importance of the contact point. Biomech Model Mechanobiol. 2007;6(3):199-210.

3. Garai K, Crick SL, Mustafi SM, Frieden C. Expression and purification of amyloid-beta peptides from Escherichia coli. Protein Expr Purif. 2009;66(1):107-12. PMCID: 2674643.

---

4. Hu X, Crick SL, Bu G, Frieden C, Pappu RV, Lee JM. Amyloid seeds formed by cellular uptake, concentration, and aggregation of the amyloid-beta peptide. Proc Natl Acad Sci U S A. 2009;106(48):20324-9. PMCID: 2787156.

5. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci U S A. 2010;107(18):8183-8. PMCID: 2889596.

6. Williamson TE, Vitalis A, Crick SL, Pappu RV. Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin. J Mol Biol. 2010;396(5):1295-309. PMCID: 2832287.

7. Mustafi SM, Garai K, Crick SL, Baban B, Frieden C. Substoichiometric inhibition of Abeta(1-40) aggregation by a tandem Abeta(40-1-Gly8-1-40) peptide. Biochem Biophys Res Commun. 2010;397(3):509-12. PMCID: 2897963.

Reviews:

1. Pappu RV, Wang X, Vitalis A, Crick SL. A polymer physics perspective on driving forces and mechanisms for protein aggregation. Arch Biochem Biophys. 2008;469(1):132-41. PMCID: 2211569.

Book Chapters:

1. Crick SL, Pappu RV. Thermodynamic and Kinetic Models for Aggregation of Intrinsically Disordered Proteins. Schweitzer-Stenner R, editor. Hoboken, NJ: John Wiley & Sons; In Press.