

**Notice 1**

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

**Notice 2**

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

## ERRATA

Pg. 9. Ln. 17: Replace “requiring” with “require.”

Pg 10. Ln 23: Delete “Protein fragments are usually more aggregation-prone than the full-length protein variants, indicating that protein regions flanking amyloidogenic sequences can suppress intermolecular interactions.”

Pg. 17. Ln 1: Replace “their” with “the” and add “of the proteins” between “insolubility” and “have”.

Pg. 19. Ln. 6: Change “ie” to “i.e.”

Pg. 27. Ln. 11: Replace “17aa” with “17 amino acid”.

Pg. 31. Ln. 18: Remove brackets between reference citations 176 and 177 and replace with comma.

Pg. 64. Ln. 4: Add “method” to the end of the sentence ending with “problems associated with this.”

Pg. 96. Ln. 7: Replace “CAG” with “glutamines”.

Pg. 144. Ln. 1: Change the first word, “Whilst”, to “Although”.

## ADDENDUM

Pg. 13. Ln 9: Comment.

The Glabe laboratory antibodies are specific for various pre-amyloidogenic conformations. These were generated against the specific species in the A $\beta$  aggregation pathway, however they are also reactive to species formed during  $\alpha$ -synuclein, IAPP, polyQ, lysozyme, insulin and prion protein aggregation, suggesting that common species are formed in the amyloidogenic pathways of different proteins.

Pg. 17. Para 2: Comment.

There is no evidence for the sidechain-sidechain interactions proposed by Perutz in the polar zipper model of polyQ aggregation.

Pg. 32: Add after Para 2.

However cellular clearance of misfolded proteins depends on a number of other factors, including substrate affinity, the rate of misfolding, and the cellular concentration of misfolded proteins. When clearance mechanisms are overtaxed by misfolded proteins, the cellular capacity to maintain protein homeostasis is compromised and this can lead to further toxicity and is associated with many diseases of aging (1).

Add after Pg. 84:

The amino acid sequence of SpA-polyQ proteins are:

MHHHHHHENLYFQGHVDVDNKFNKEQQNAFWELHLPNLNEEQRNQFIQSLKDDPSQSANLLAEAKKLNDA  
QAPKPR(Q)<sub>n</sub>PG

(Q)<sub>n</sub> refers to the polyQ tract of specific repeat length (n).

Pg. 78: Add after Results Paragraph 1 of inserted manuscript.

The CD analysis does not take into account the linking sequence, Hexahis tag and TEV site. The conformations of these regions are unknown and may contribute to either the alpha-helical or random coil spectral contribution. The main message from the CD data is that there is an increase in random coil content with increasing polyQ length. Although the increase in % random coil contribution may not be entirely consistent with the increasing polyQ tract length, this could be attributed to the CD technique and spectral deconvolution algorithm.

Pg 81. Para 1: Comment.

The high intensity crosspeaks in the NMR spectra at 8.4, 7.6 and 6.9 ppm are consistent with the chemical shifts of backbone and sidechain glutamine amide groups when glutamine is in a random coil conformation (2). These peaks would be shifted upfield or downfield relative to the random coil values if the glutamines were in an  $\alpha$ -helical or  $\beta$ -extended conformation respectively. The intensity of these peaks is consistent with polyQ repeat length, i.e. these peaks are much lower in intensity in the SpA-cQ5 spectra as compared to the SpA-cQ52 spectra further confirming that these peaks arise due to the presence of the polyQ tract.

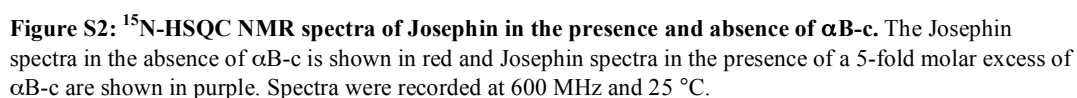
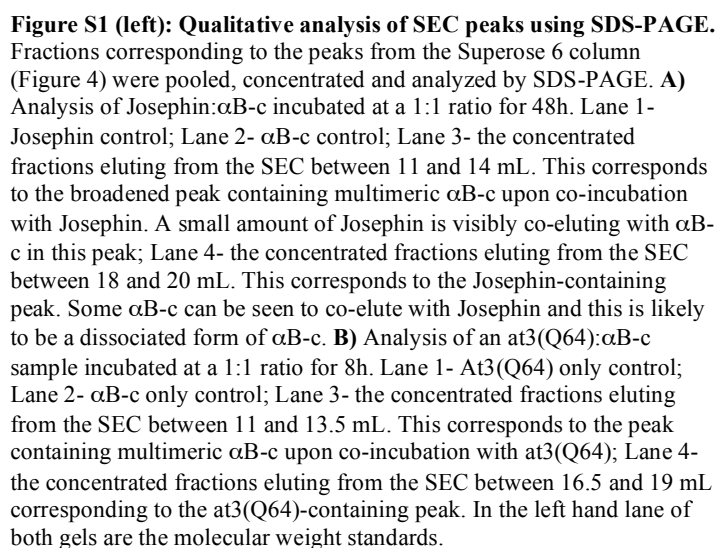
Pg 99. Ln 5: Add at the end of sentence.

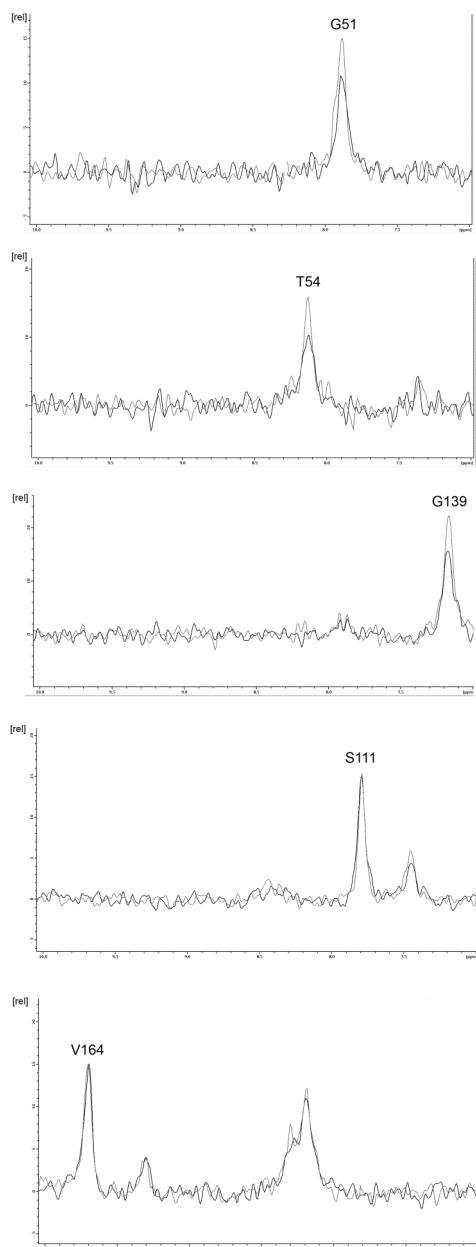
Protein concentration was determined by measuring the absorbance at 280nm. Protein concentration was independently verified by unfolding the proteins in GuSCN and determining the fluorescence intensity at  $\lambda=355$  nm ( $\lambda_{ex}=280$  nm), and comparing the intensity relative to a standard of known concentration.

These statistical analyses, however, may be somewhat biased due to the small size of the disease polyQ data set, due to the fact that there are only nine polyQ disease proteins.

The numbers outside of the brackets refer to the actual number of proteins in each group. The numbers in brackets represent the percentage of all, non-disease or disease proteins with a polyQ tract in each position. For example, 29% of all polyQ-containing proteins have an N-terminal polyQ tract.

Insert the additional data as Supplementary figures S1, S2 and S3 for Chapter 6 after Pg. 141.





**Figure S3 (left):  $\alpha$ B-c partially shields specific residues from paramagnetic relaxation enhancement.**  $^1\text{H}$  slices through  $^{15}\text{N}$ -HSQC spectra of  $^{15}\text{N}$ -Josephin in buffer containing the paramagnetic relaxation enhancement agent 5 mM Gd(DTPA-BMA). In black is shown apo  $^{15}\text{N}$ -josephin and in grey is  $^{15}\text{N}$ -Josephin in the presence of equimolar unlabelled  $\alpha$ B-c. Residues G51, T54 and G139 are partially shielded from paramagnetic relaxation enhancement by the presence of  $\alpha$ B-c, whereas residues S111 and V164 are not.

Pg. 147. Para 2: Add “In our SpA-polyQ model system (Chapter 4), the composition of the immediately flanking regions is unchanged amongst the proteins, therefore the polyQ biophysical properties due to local sequence is consistent amongst the proteins.” after the sentence, “The composition of regions flanking repeats can alter the biophysical behaviour of the polyQ region and adjust the kinetic barrier to amyloid formation.”

## References

1. Balch, W. E., R. I. Morimoto, A. Dillin, and J. W. Kelly. (2008) Adapting proteostasis for disease intervention. *Science* 319:916-919.
2. Wuthrich K (1986). “The Foundations: Structure and NMR of Biopolymers,” in *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, USA. Pg. 13-25.



---

# **An investigation into polyglutamine aggregation mechanisms**

Amy Robertson, BBiomedSci. (Hons)

Submitted for the Degree of Doctor of Philosophy

December 2009



**TABLE OF CONTENTS**

|   |               |
|---|---------------|
| <b>Table of Contents</b>  | <b>i</b>      |
| <b>Acknowledgements</b>   | <b>iv</b>     |
| <b>General Declaration</b>  | <b>v</b>      |
| <b>Declaration of Authorship for Chapter 3</b>  | <b>vi</b>     |
| <b>Declaration of Authorship for Chapter 4</b>  | <b>vii</b>    |
| <b>Declaration of Authorship for Chapter 5</b>  | <b>ix</b>     |
| <b>Declaration of Authorship for Chapter 6</b>  | <b>x</b>      |
| <b>Abbreviations</b>  | <b>xii</b>    |
| <b>Abstract</b>   | <b>xiv</b>    |
| <br><b><u>Chapter 1: Introduction</u></b>   | <br><b>1</b>  |
| <b>1.1 Protein Misfolding and Aggregation</b>   | <b>2</b>      |
| 1.1.1 Protein Conformational Diseases   | 2             |
| 1.1.2 Amyloid Fibrils   | 5             |
| 1.1.3 Mechanisms of Aggregation   | 8             |
| <i>1.1.3.1 Initiation of aggregation from the monomeric state</i>                     | <i>9</i>      |
| <i>1.1.3.2 Kinetics of aggregation</i>  | <i>11</i>     |
| <i>1.1.3.3 Amyloid oligomers and toxicity</i>   | <i>12</i>     |
| <b>1.2 Polyglutamine Misfolding</b>   | <b>14</b>     |
| 1.2.1 Polyglutamine Diseases  | 14            |
| 1.2.2 The Physical Behaviour and Aggregation of PolyQ                                 | 17            |
| 1.2.3 The Protein Context of PolyQ  | 20            |
| 1.2.4 PolyQ Expansion and Protein Stability   | 22            |
| 1.2.5 Proteolytic Cleavage: A Protective Role of Protein Context                      | 25            |
| 1.2.6 Multidomain Aggregation: Host Protein Domains that Assist PolyQ Aggregation     | 26            |
| 1.2.7 Modulation of Misfolding, Aggregation and Toxicity through Interaction Networks | 30            |
| <b>1.3 Aims of this Project</b>   | <b>34</b>     |
| <br><b><u>Chapter 2: Materials and Methods</u></b>                                    | <br><b>37</b> |
| <b>2.1 Reagents</b>   | <b>38</b>     |
| <b>2.2 Microbial Techniques</b>   | <b>38</b>     |
| 2.2.1 Growth Media  | 38            |
| 2.2.2 Bacterial Strains   | 39            |
| 2.2.3 Oligonucleotide Design  | 39            |
| 2.2.4 Polymerase Chain Reaction   | 40            |
| 2.2.5 QuikChange™ site-directed Mutagenesis   | 41            |
| 2.2.6 Agarose gel Electrophoresis   | 42            |
| 2.2.7 Restriction Enzyme Digestion  | 42            |
| 2.2.8 Gel Purification  | 43            |
| 2.2.9 Ligations   | 43            |
| 2.2.10 Plasmid Purification   | 43            |

|  |                  |
|--|------------------|
| 2.2.11 Preparation of Competent Cells.....   | 43               |
| 2.2.12 Bacterial Transformations.....  | 44               |
| 2.2.13 Colony Screening.....   | 44               |
| <b>2.3 Preparation of Proteins.....</b>  | <b>46</b>        |
| 2.3.1 Buffers/Solutions.....   | 46               |
| 2.3.2 Expression of SpA-polyQ variants.....  | 46               |
| 2.3.3 Expression of Ataxin-3 variants.....   | 46               |
| 2.3.4 Expression of <sup>15</sup> N and <sup>15</sup> N/ <sup>13</sup> C-labeled proteins for NMR spectroscopy.....          | 47               |
| 2.3.5 Purification of SpA-polyQ proteins.....  | 48               |
| 2.3.6 Purification of Ataxin-3 variants.....   | 48               |
| 2.3.7 Preparation of αB-crystallin.....  | 49               |
| 2.3.8 Determination of Protein Concentration.....  | 49               |
| <b>2.4 SDS-PAGE.....</b>   | <b>50</b>        |
| <b>2.5 Size Exclusion Chromatography.....</b>  | <b>51</b>        |
| <b>2.6 Circular Dichroism.....</b>   | <b>52</b>        |
| <b>2.7 Thermal and Guanidine Denaturation.....</b>   | <b>53</b>        |
| 2.7.1 Thermal Denaturation.....  | 53               |
| 2.7.2 GuHCl unfolding.....   | 53               |
| <b>2.8 Protein Aggregation assays.....</b>   | <b>54</b>        |
| 2.8.1 SpA-polyQ variants.....  | 54               |
| 2.8.2 Ataxin-3 variants.....   | 54               |
| 2.8.3 Aggregation assays with the addition of αB-c.....  | 54               |
| 2.8.4 QBP1 assays.....   | 55               |
| 2.8.5 ThioflavinT fluorescence.....  | 55               |
| 2.8.5.1 Discontinuous ThT measurements.....  | 55               |
| 2.8.5.2 Continuous ThT measurements.....   | 55               |
| <b>2.9 Membrane Filter Trap Assay.....</b>   | <b>56</b>        |
| <b>2.10 Transmission Electron Microscopy.....</b>  | <b>57</b>        |
| <b>2.11 Nuclear Magnetic Resonance (NMR) Spectroscopy.....</b>   | <b>57</b>        |
| <b><u>Chapter 3: A cyclic method for the incremental expansion of polyglutamine repeats in recombinant proteins.....</u></b> | <b><u>59</u></b> |
| <b><u>Addendum to Chapter 3: Materials and Methods.....</u></b>  | <b><u>71</u></b> |
| <b><u>Chapter 4: The structural impact of a polyglutamine tract is location-dependent.....</u></b>                           | <b><u>75</u></b> |

|   |            |
|---|------------|
| <b><u>Chapter 5: Protein domain architecture determines the aggregation potential of polyglutamine repeats</u></b> .....          | <b>85</b>  |
| <b><u>Chapter 6: Small heat-shock proteins inhibit polyglutamine aggregation by interactions with a flanking domain</u></b> ..... | <b>109</b> |
| <b><u>Chapter 7: General Discussion</u></b> .....   | <b>143</b> |
| <b>7.1 Pathways to PolyQ Aggregation</b> .....  | <b>144</b> |
| 7.11 PolyQ Self-association.....  | 144        |
| 7.12 Influence of Flanking domains on PolyQ Aggregation.....  | 145        |
| 7.13 Small heat shock proteins and PolyQ Aggregation.....   | 149        |
| <b>7.2 Native State Energetics and the Role of Flanking Domains</b> .....   | <b>151</b> |
| <b>7.3 Further questions about PolyQ Misfolding</b> .....   | <b>152</b> |
| <b><u>Appendix 1: Polyglutamine fibril architecture is influenced by glutamine tract length and flanking regions</u></b> .....    | <b>155</b> |
| <b><u>References</u></b> .....  | <b>184</b> |

## **Acknowledgements**

Over the past four years I have been very fortunate to work with a great group of scientifically skilled and lovely people. The work of this thesis would not have been possible without the ideas, collaboration and support of many people.

I would not have embarked upon and continued this project if it were not for the enthusiasm, ideas, support and guidance of my supervisor, Steve Bottomley. I thank him for not always thinking my ideas are crazy and for providing an environment that has given me the freedom and confidence to work independently.

This project has grown from difficult beginnings, and I would like to thank Lisa Cabrita for teaching me the cloning skills that I have utilized consistently. I have benefited greatly from working with a very helpful group of people on polyQ projects, including Helen, Bronwen, Andy and Su. I would also like to thank all of the lovely people I have worked with in the Bottomley lab, especially the other PhD students with whom I have shared this experience, including Helen, Anja, Vita, Su and Austin. Also, I would like to thank Glyn and Mary for their reading and criticism of this thesis.

I am very grateful to a number of collaborators. Martin Scanlon, James Horne and Steve Headey have helped to perform the NMR spectroscopy in this thesis. I would like to thank John Carver for the initial idea to perform the experiments with  $\alpha$ -crystallin, and Heath Ecroyd for giving advice on these and providing protein. I would also like to thank Ashley Buckle and Mark Bate for helping to perform the bioinformatic analysis of polyQ proteins.

I have been extremely lucky to be supported and entertained by some wonderful people outside of the lab during my PhD. Big thanks to my housemates and friends, Kathryn and Clare, and to the rest of my great friends, especially Anja, Joanna and Cathy.

Most of all I would like to thank my family. I am forever indebted to my Mum, Dad and Sarah for all of their love and support which have made my PhD possible.

## General Declaration

In accordance with Monash University Doctorate Regulation 17/ Doctor of Philosophy and Master of Philosophy (MPhil) regulations the following declarations are made:

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes one original paper published in a peer reviewed journal and three unpublished publications. The core theme of the thesis is polyglutamine misfolding and aggregation. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Department of Biochemistry and Molecular Biology under the supervision of Professor Stephen P. Bottomley.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of Chapters 3, 4, 5, and 6 my contribution to the work involved the following:

| Thesis chapter | Publication title  | Publication status | Nature and extent of candidate's contribution  |
|----------------|--|--------------------|--|
| 3              | A cyclic method for the incremental expansion of polyglutamine repeats in recombinant proteins     | submitted          | Idea for and design of method, all laboratory work, manuscript preparation   |
| 4              | The structural impact of a polyglutamine tract is location-dependent                               | published          | Experimental design, protein construction and preparation, protein characterization, folding studies, contribution to NMR studies, manuscript preparation. |
| 5              | Protein domain architecture determines the aggregation-potential of polyglutamine repeats          | submitted          | Protein construction and preparation, experimental work, intellectual input for bioinformatic analysis, manuscript preparation                             |
| 6              | Small heat-shock proteins inhibit polyglutamine aggregation by interactions with a flanking domain | submitted          | Protein preparation, performed experiments, prepared manuscript.   |

I have renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

**Signed:** .....

**Date:** .....

## Declaration of Authorship for Chapter 3

In the case of Chapter 3, the following co-authors contributed to the work:

| Name              | Nature of contribution   | % contribution for student co-authors |
|-------------------|--|---------------------------------------|
| Amy Robertson     | Idea for and design of method, all laboratory work, manuscript preparation | 90                                    |
| Lisa Cabrita      | Intellectual input, manuscript preparation                                 |                                       |
| Stephen Bottomley | Supervision, intellectual input, manuscript preparation                    |                                       |

Candidate's

Signature: .....

*Amy Robertson*

Date: .....

### Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

|                  |   |
|------------------|---|
| <b>Location:</b> | Department of Biochemistry and Molecular Biology, Monash University |
|------------------|---|

Signature:

.....

*Lisa Cabrita*

Date: .....17/11/09...

Signature:

.....

*Stephen Bottomley*

Date: .....9/12/09.....



## Declaration of Authorship for Chapter 4

In the case of Chapter 4, the following co-authors contributed to the work:

| Name              | Nature of contribution   | % contribution for student co-authors |
|-------------------|--|---------------------------------------|
| Amy Robertson     | Experimental design, protein construction and preparation, protein characterization, folding studies, contribution to NMR studies, manuscript preparation. | 60                                    |
| James Horne       | NMR experiments and structure determination.   |                                       |
| Andrew Ellisdon   | Intellectual input, manuscript preparation.  |                                       |
| Bronwen Thomas    | Aggregation experiments.   | 10                                    |
| Martin Scanlon    | Intellectual input, manuscript preparation.  |                                       |
| Stephen Bottomley | Supervision, intellectual input, manuscript preparation.   |                                       |

Candidate's


Signature: ..... Date: .....  
*Amy Robertson*


### Declaration by co-authors


The undersigned hereby certify that:


1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

|                   |   |
|-------------------|---|
| <b>Locations:</b> | Department of Biochemistry and Molecular Biology, Monash University<br>Monash Institute of Pharmaceutical Sciences, Monash University |
|-------------------|---|

Signature: .....  ..... Date: 06/11/2009  
*James Horne*

Signature: .....  ..... Date: 3/11/09  
*Andrew Ellisdon*

Signature: .....  ..... Date: 4/11/09  
*Bronwen Thomas*

Signature: .....  ..... Date: 17/11/09  
*Martin Scanlon*

Signature: .....  ..... Date: 9/12/09  
*Stephen Bottomley*

## Declaration of Authorship for Chapter 5

In the case of Chapter 5, the following co-authors contributed to the work:

| Name              | Nature of contribution   | % contribution for student co-authors |
|-------------------|--|---------------------------------------|
| Amy Robertson     | Protein construction and preparation, experimental work, intellectual input for bioinformatic analysis, manuscript preparation | 80                                    |
| Mark Bate         | Computational work for bioinformatics analysis   |                                       |
| Ashley Buckle     | Intellectual input   |                                       |
| Stephen Bottomley | Supervision, intellectual input, manuscript preparation  |                                       |

Candidate's

Signature: ..... Date: .....  
*Amy Robertson*

### Declaration by co-authors

The undersigned hereby certify that:

1. the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. there are no other authors of the publication according to these criteria;
5. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

|                  |   |
|------------------|---|
| <b>Location:</b> | Department of Biochemistry and Molecular Biology, Monash University |
|------------------|---|

Signature: ..... Date: *9/12/09*.....  
*Mark Bate*

Signature: ..... Date: *8/12/09*.....  
*Ashley Buckle*

Signature: ..... Date: *9/12/09*.....  
*Stephen Bottomley*

## Declaration of Authorship for Chapter 6

### Declaration by candidate

In the case of Chapter 6, the following co-authors contributed to the work:

| Name              | Nature of contribution  | % contribution for student co-authors |
|-------------------|---|---------------------------------------|
| Amy Robertson     | Protein preparation, performed experiments, prepared manuscript.  | 70                                    |
| Stephen Headey    | NMR experiments, manuscript preparation.                          |                                       |
| Helen Saunders    | Electron microscopy, intellectual input                           | 10                                    |
| Heath Ecroyd      | Intellectual input, supplied $\alpha$ B-c, manuscript preparation |                                       |
| John Carver       | Intellectual input, supplied $\alpha$ B-c, manuscript preparation |                                       |
| Martin Scanlon    | Intellectual input  |                                       |
| Stephen Bottomley | Supervision, intellectual input, manuscript preparation           |                                       |

Candidate's

Signature: ..... Date: .....


*Amy Robertson*

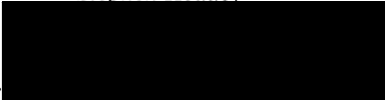
### Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

|                   |   |
|-------------------|---|
| <b>Locations:</b> | Department of Biochemistry and Molecular Biology, Monash University<br>Monash Institute of Pharmaceutical Sciences, Monash University |
|-------------------|---|

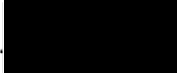
Signature: .....  ..... Date: 8/12/09  
*Stephen Headey*

Signature: .....  ..... Date: 8/12/09  
*Helen Saunders*

Signature: .....  ..... Date: 5/11/09  
*Heath Ecroyd*

Signature: .....  ..... Date: 5/11/09  
*John Carver*

Signature: .....  ..... Date: 12/11/09  
*Martin Scanlon*

Signature: .....  ..... Date: 9/12/09  
*Stephen Bottomley*

**ABBREVIATIONS**

|                    |  |
|--------------------|--|
| A $\beta$          | Amyloid- $\beta$                         |
| Abs <sub>280</sub> | Absorbance at 280 nm                     |
| AFM                | Atomic Force Microscopy                  |
| Amp                | Ampicillin                               |
| APS                | Ammonium Persulfate                      |
| AR                 | Androgen receptor                        |
| At3                | Ataxin-3                                 |
| At3(Q64)           | Ataxin-3 with a 64 glutamine repeat      |
| $\beta_2$ M        | $\beta_2$ -microglobulin                 |
| BCA                | Bicinchoninic Acid                       |
| BME                | $\beta$ -mercaptoethanol                 |
| BSA                | Bovine Serum Albumin                     |
| CACNA1A            | alpha1 voltage-dependent calcium channel |
| CI2                | Chymotrypsin Inhibitor 2                 |
| CD                 | Circular dichroism                       |
| CRABP1             | Cellular retinoic acid binding protein 1 |
| DMSO               | Dimethyl Sulfoxide                       |
| DRPLA              | Dentatorubral-pallidoluysian atrophy     |
| DTT                | Dithiothreitol                           |
| ECL                | Enzymatic chemiluminescence              |
| EDTA               | Ethylenediaminetetraacetic Acid          |
| EPR                | Electron paramagnetic resonance          |
| FPLC               | Fast Protein Liquid Chromatography       |
| FRET               | Fluorescence resonance energy transfer   |
| FTIR               | Fourier transform infrared               |
| GST                | Glutathione S-transferase                |
| GuHCl              | Guanidium Hydrochloride                  |
| GuSCN              | Guanidium Thiocyanate                    |
| HD                 | Huntington's disease                     |
| Hsp                | Heat shock protein                       |
| HSQC               | Heteronuclear Single Quantum Correlation |
| Htt                | Huntingtin                               |
| Ig                 | Immunoglobulin                           |

---

|                   |   |
|-------------------|---|
| IPTG              | Isopropyl- $\beta$ -D-thiogalactopyranoside     |
| $\lambda_{em}$    | Emission Wavelength                             |
| $\lambda_{ex}$    | Excitation Wavelength                           |
| Mb                | Myoglobin                                       |
| MJD               | Machado Joseph disease                          |
| NMR               | Nuclear Magnetic Resonance                      |
| NOESY             | Nuclear Overhauser Effect Spectroscopy          |
| OD <sub>600</sub> | Optical Density at 600 nm                       |
| OPMD              | Oculopharyngeal muscular dystrophy              |
| QBP1              | PolyQ-binding peptide 1                         |
| PABPN1            | Polyadenine binding protein 1                   |
| PAGE              | Polyacrylamide Gel Electrophoresis              |
| PCR               | Polymerase Chain Reaction                       |
| PMSF              | Phenylmethyl sulphonyl fluoride                 |
| PolyA             | Polyalanine                                     |
| PolyP             | Polyproline                                     |
| PolyQ             | Polyglutamine                                   |
| PRE               | Paramagnetic Relaxation Enhancement             |
| SBMA              | Spinal bulbar muscular atrophy                  |
| SCA               | Spinocerebellar ataxia                          |
| SDS               | Sodium dodecyl sulphate                         |
| SEC               | Size Exclusion Chromatography                   |
| sHsp              | small heat shock protein                        |
| SpA               | <i>Staphylococcus aureus</i> Protein A B domain |
| SpA-cQn           | SpA with C-terminal polyQ                       |
| SpA-iQn           | SpA with intradomain polyQ                      |
| SpA-idQn          | 2 SpA domains with a polyQ tract between them   |
| $\theta$          | Molar Ellipticity                               |
| TBP               | TATA-binding protein                            |
| TBS               | Tris-buffered saline                            |
| TEM               | Transmission Electron Microscopy                |
| TEMED             | Tetramethylethylenediamine                      |
| ThT               | Thioflavin T                                    |
| TRiC              | TCP-1 ring complex                              |
| UV                | Ultraviolet                                     |

**ABSTRACT**

Protein aggregation is a key mechanism involved in neurodegeneration associated with Alzheimer's, Parkinson's and Huntington's diseases. Nine diseases (including Huntington's) arise from polyglutamine (polyQ) expansion above a repeat threshold of approximately 37 glutamines, and neuronal toxicity correlates with the process of protein aggregation. The similar toxic gain-of-function mechanism of the nine diseases supports the hypothesis that disease onset and progression is dependent upon polyQ expansion. However, there is an increasing body of literature demonstrating that the protein context of the polyQ tract plays an important modulating role in the disease process. The composition of regions flanking repeats can alter the biochemical and biophysical properties of the polyQ region. Interactions between flanking regions and other molecules can also influence aggregation and cellular localization, which are critical factors for toxicity. More recently, there is evidence that domains flanking the polyQ tract can also aggregate independently of the polyQ tract, and that this significantly alters the rate at which the polyQ regions form fibrillar aggregates and the properties of these aggregates. This thesis investigates the molecular mechanisms leading to polyQ aggregation and the role of protein context in modulating the aggregation pathway. A family of model polyQ proteins were engineered and produced. The proteins have a repeat-length dependent aggregation mechanism, recapitulating the relationship between repeat length and age of disease onset. The stability and structure of the flanking domain were unperturbed when fused to a pathological length polyQ tract, suggesting that protein misfolding within the polyQ tract is the driving force behind the key characteristics of the polyQ diseases. The repeat location and domain architecture affect the rate of polyQ-dependent aggregation, indicating that host protein factors can modulate aggregation. Furthermore, the small heat shock protein,  $\alpha$ B-



---

crystallin does not inhibit the aggregation of the model polyQ protein that aggregates by the polyQ-driven mechanism.  $\alpha$ B-crystallin does, however, inhibit aggregation of the disease protein ataxin-3 by a mechanism involving interactions with the flanking domain. Together, the results within this thesis have provided insight into the molecular basis of polyQ disease and have shown that the propensity for polyQ aggregation is determined by a complex interplay between the polyQ region, host protein factors and the cellular environment.



# **Chapter 1**

## **Introduction**

## **1.1 Protein Misfolding and Aggregation**

### **1.1.1 Protein Conformational Diseases**

The ability of a protein to attain and maintain its correct native fold is essential for functional activity and cellular homeostasis. When proteins cannot fold correctly they can misfold and aggregate. Genetic and/or environmental changes can cause proteins to be susceptible to misfolding and aggregation, and over 40 human diseases have an identified link to protein aggregation (1), broadly classed as ‘conformational diseases’. The protein aggregates formed are generally resistant to degradation resulting in their cellular or extracellular accumulation as inclusions or plaques. This group of diseases has variable phenotypes, ranging from neurodegenerative diseases to systemic amyloidoses, however, they share similar underlying molecular mechanisms that cause the proteins to form ordered, amyloid-like aggregates that are fibrillar and  $\beta$ -sheet rich (Table 1). Within this thesis, all references to amyloid and fibrils refer to amyloid-like fibrils.

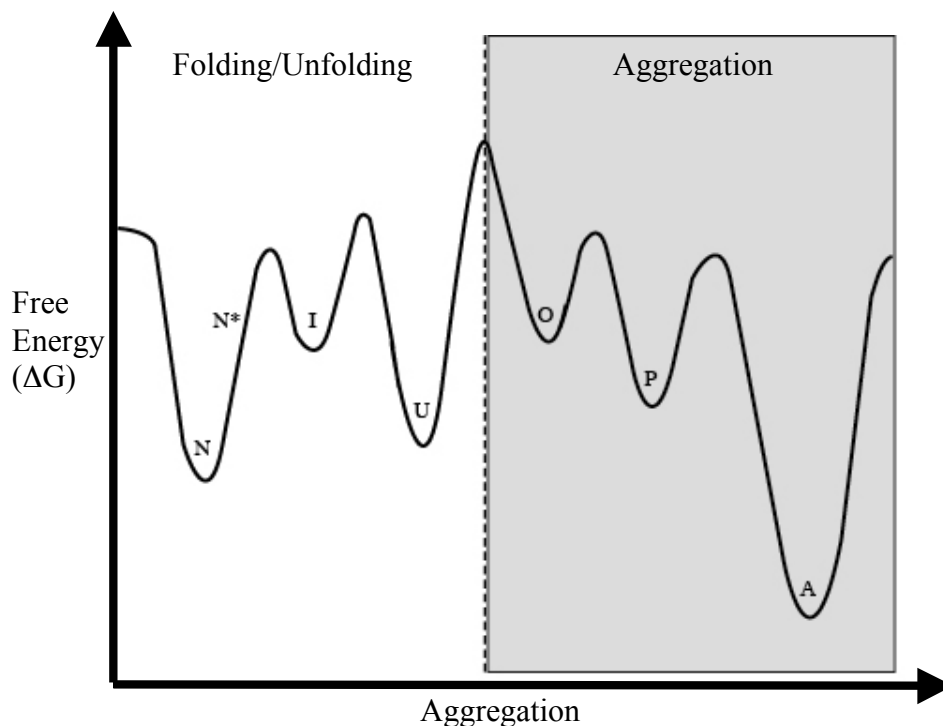
**Table 1: Proteins that initiate some of the conformational diseases**

| <b>Disease</b>                         | <b>Aggregating protein or peptide</b> |
|--|---------------------------------------|
| Alzheimer’s disease                    | Amyloid $\beta$ peptide               |
| Parkinson’s disease                    | $\alpha$ -synuclein                   |
| Spongiform encephalopathies            | Prion protein/fragments               |
| Amyotrophic lateral sclerosis          | Superoxide dismutase 1                |
| Huntington’s disease                   | Huntingtin fragments (polyQ)          |
| Spinocerebellar ataxia type 3          | Ataxin-3 (polyQ)                      |
| Lysozyme amyloidosis                   | Mutants of lysozyme                   |
| Haemodialysis-related amyloidosis      | $\beta_2$ -microglobulin              |
| Type II diabetes                       | Islet amyloid polypeptide             |
| Familial amyloidotic polyneuropathy    | Mutants of transthyretin              |
| Hereditary Cerebral amyloid angiopathy | Mutants of cystatin C                 |

A protein's energy landscape consists of the maximal number of conformational states accessible to a given polypeptide sequence (Figure 1.1). Protein aggregation challenges the fundamental tenet of protein folding, Anfinsen's hypothesis, that the polypeptide sequence determines a unique native structure of lowest free energy (2). Experimental and theoretical studies suggest that amyloid fibrils can represent a lower free energy state than the native conformation, and are in fact accessible by all proteins, however, their formation is limited by a large kinetic barrier (3, 4). Perturbations to the thermodynamic and kinetic parameters favouring the native state by mutation, altered pH, ionic concentration and/or altered physiological interactions and chaperone networks can lower the barrier leading to aggregation (3, 4). Under physiological conditions, the difference in free energy between the native and unfolded states is, on average, 5-15 kcal/mol, therefore the native state is only moderately stable (4). Therefore small changes can significantly alter the equilibrium between different conformations.

Much of the current knowledge of protein misfolding has been derived from isolated reactions within dilute solutions, however, in the cell, collisions between molecules under different conditions increase the complexity of the energy landscape. Many amyloid diseases are associated with ageing, and there is considerable evidence that proteostasis networks have an important role in disease onset and progression (5). Proteostasis refers to the cellular properties that control protein conformation and interactions (5). Molecular chaperones and cellular degradative proteins are functionally important in the maintenance of proteostasis, recognizing misfolded substrates leading to their refolding or degradation. However, the functions of these systems are compromised as cells age. Chaperones and proteasomal components associate with intracellular inclusions, indicating their *in vivo* importance and tight link with disease.

Preventing the aberrant protein conformational changes leading to the formation of amyloid-like aggregates is key to avoiding aggregation. A large breadth of research has focused on the properties of endpoint amyloid aggregates and the mechanisms by which they are formed. These will be discussed in the first part of this chapter.



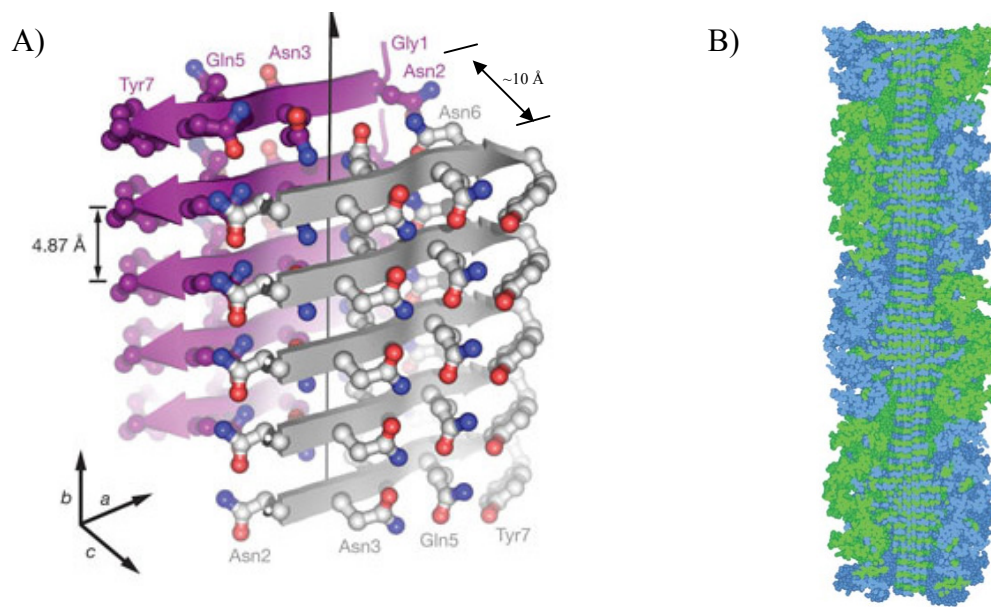
**Figure 1.1: Free energy landscape of folding/unfolding and aggregation.** Local conformational fluctuations around the native state (N), lead to the formation of near-native ( $N^*$ ) states. Native state destabilization can also lead to the population of partially-unfolded intermediates (I). Aggregation can proceed from  $N^*$ , I and unfolded (U) states. These conformations must traverse the high kinetic barrier that leads to the aggregation side of the landscape. As aggregation is dependent on intermolecular interactions, the height of the kinetic barrier is protein concentration-dependent. Alternative pathways can lead to the transient population of oligomers (O), protofibrils (P) and the highly stable minimal energy amyloid (A) state.

### **1.1.2 Amyloid Fibrils**

Amyloid fibrils are proposed to be the lowest energy conformational state for all proteins (Figure 1.1). It is therefore essential to determine the properties and structures of amyloid in order to understand disease-related misfolding. Fibrils are typically protease-resistant and detergent-insoluble (6, 7). They interact with the fluorescent probes Congo-red and ThioflavinT, and specifically bind a number of amyloid-binding antibodies (8-10). Despite large variability in sidechain topologies, which give rise to different substructures, amyloid fibrils appear to have some generic structural properties (4, 6, 8). Morphologically, fibrils are unbranched and range between 5-15 nm in width when observed by electron microscopy (11).

Structural models for amyloid fibrils have been built from X-ray fibre diffraction data, suggesting that fibrils have a common core molecular architecture (12-14). Fibrils have a signature diffraction pattern, showing two characteristic signals; a sharp meridional reflection at 4.7 Angstroms perpendicular to the fibril axis, indicative of the interstrand distance, and a broad equatorial reflection at around 10 Angstroms, indicating the intersheet distance (14, 15). This diffraction pattern suggests that fibrils have a cross- $\beta$  secondary structure, in which  $\beta$ -strands that form  $\beta$ -sheets in the fibrils run perpendicular to the long axis of the fibril with hydrogen bonds running parallel to the fibril axis to link the strands (Figure 1.2 A). Fibrils are therefore highly stable due to the extensive hydrogen-bonding network.

Amyloid fibrils are not crystalline, limiting the ability to obtain crystallographic structural data. Some detail on intermolecular contacts has been obtained from X-ray diffraction of several different amyloidogenic peptides associating to form microcrystals (16, 17). Within the microcrystals, all of the peptides have a cross- $\beta$  arrangement. Fibrils formed from the prion-derived GNNQQNY peptide form



**Figure 1.2: Amyloid fibril structure.**

A) Cross- $\beta$  core architecture of amyloid as determined for the GNNQQNY peptide by X-ray crystallography. The purple and grey represent two  $\beta$ -sheets. Within each  $\beta$ -sheet, the strands are in a parallel arrangement. The interstrand distance is shown as 4.87 Angstrom, and the intersheet distance is approximately 10 Angstrom. Peptides can also exist in alternative arrangements. (Adapted from ref. (16)) B) Representative figure of an amyloid cross- $\beta$  spine surrounded by globular domains. (From ref. (18))

$\beta$ -strands aligned in a parallel arrangement within each sheet, and the opposing sheets are aligned antiparallel to each other (16). While the in-register parallel arrangement is the most common amyloid arrangement observed in structural studies, there is evidence that the precise interstrand arrangement varies depending on the peptide sequence (19). Some peptides form fibrils with an antiparallel arrangement of  $\beta$ -strands, and some are thought to form a  $\beta$ -helical structure (17, 20, 21).

Protein size limitations have hindered solution NMR studies of fibril structure, however, several groups have utilized solid state NMR to probe the structure of these species (22). Models of a transthyretin peptide and A $\beta$  fibrils were consistent with the crystallographic data, suggesting that the fibril core was in a parallel cross- $\beta$  conformation (23, 24), and this is consistent with models obtained from site-directed



spin-labeling and EPR spectroscopy (19). The A $\beta$  fibril model also included regions flanking the fibril core and indicated that the adjacent peptide sequence was in a disordered conformation, demonstrating that not all of the residues in a fibril adopt  $\beta$ -sheet structure (25).

Polymorphisms in fibril structure are determined by growth conditions, polypeptide length and side-chain topology (19). Extrinsic conditions, such as solution pH, ionic concentration, temperature and agitation can affect fibril morphology (26). Peptide length correlates with the orientation of  $\beta$ -strands; short peptides favour an antiparallel orientation whilst longer peptides favour a parallel in-register orientation. The specific sidechains determine the intersheet distance, which can range from 8-12 Angstroms depending on the sidechain volume and the favourability of sidechain packing (12). Fibril stability is influenced by the hydrogen-bonding networks of stacked strands which are determined by the amino acid composition.

Different sidechains can stabilize/destabilize amyloid fibrils, and/or promote/inhibit their formation. Many proteins contain relatively small aggregation-prone segments. Regions of high aggregation propensity often correlate with regions prone to  $\beta$ -strand formation and can be predicted by a number of algorithms, including TANGO (27), Zyggregator (28) and BETASCAN (29), which predict aggregation-prone sequences based upon parameters including hydrophobicity, electrostatic interactions and hydrogen-bonding (27-29). Fibril core regions are rich in hydrophobic residues due to the favourability of side-chain burial (19), whilst charged residues are generally destabilizing due to electrostatic repulsion (30, 31). Aromatic residues can stabilize fibrils by  $\pi$ -bonding between adjacent strands (32) and prolines are unfavourable in the fibril core as they are  $\beta$ -sheet breakers (33). Homopolymeric amino acid repeats, such as polyglutamine and polyalanine, are intrinsically capable of

forming fibrillar aggregates, which could be attributed to high local charge and/or hydrophobicity (34, 35).

In many cases, aggregates are composed of full-length proteins with only part of the polypeptide sequence involved in the cross- $\beta$  core, therefore, native-like protein regions can flank the fibril core, influencing morphology and functional properties (Figure 1.2 B). Domains of the yeast prions Ure2p and Sup35p maintain their native conformation in the fibril context (36, 37) and domains flanking core polyalanine sequences can retain their folded conformation (38). These are all multidomain proteins in which the amyloidogenic sequence is outside the boundaries of the folded domain. A variant of RNase A with an engineered intradomain amyloid-competent sequence forms fibrils in which a single  $\beta$ -strand from each monomeric subunit is contributed to the fibril core by a domain-swapping mechanism (18). This chimeric protein retains the native activity of RNase A. There is evidence that a similar domain swapping mechanism is involved in cystatin C aggregation (39). These examples indicate that domains can retain their folded, functional conformation within the fibril, and essentially decorate the stable fibril core. The strength and rigidity of cross- $\beta$  fibrils and the potential of fusing functional peptide regions has been taken advantage of by nanotechnologists to prepare rigid yet functional biomaterials (40).

### **1.1.3 Mechanisms of Aggregation**

Fibril formation is initiated by perturbations to the native conformation. After initiation, aggregation occurs via an ordered process, often involving the formation of small oligomeric and fibrillar aggregates which are thought to be the species involved in toxicity.

---

**1.1.3.1 Initiation of aggregation from the monomeric state**

The initial molecular events in the aggregation cascades of most proteins occur within the monomer. This raises the question of what factors enhance the favourability of intermolecular over intramolecular interactions? There are several mechanisms by which aggregation can be initiated from monomeric native conformations.

Proteins are rich in amyloidogenic motifs, as identified by various algorithms that predict aggregation-prone regions based upon primary sequence (27, 33, 41). However, burial of these motifs within the tertiary structure may shield and conformationally restrict amyloid-prone regions such that the aggregation-competent conformation cannot be sampled (42, 43). Protein conformational changes, leading to global destabilization, altered local dynamics and/or proteolytic cleavage can lead to enhanced exposure of amyloidogenic regions (43).

Native state destabilization reduces the intramolecular interactions that favour maintenance of the native state, in favour of intermolecular interactions that lead to  $\beta$ -sheet and amyloid formation. Native lysozyme is mostly  $\alpha$ -helical and some mutants partially unfold prior to forming  $\beta$ -sheet rich aggregates (44, 45). Similarly,  $\beta_2$ -microglobulin ( $\beta_2$ M) aggregates under acidic conditions requiring partial unfolding (46). Partially-denaturing conditions can cause numerous other non-disease related proteins to form amyloid-like aggregates and oligomeric complexes (47-49). These studies suggest that amyloid formation is a generic property of polypeptides under destabilizing solution conditions.

Highly dynamic proteins undergo random thermal fluctuations around the minimal energy native state leading to the formation of locally unfolded conformations. Local fluctuations occur at a more rapid rate than global unfolding, leading to conformations that do not cross the major unfolding energy barrier, rather, such

conformations are described as near-native, as they retain mostly native structure (42). Mutations may alter the kinetics of the conformational fluctuations leading to local unfolding and enhance the population of amyloid-competent states (42, 44, 50). Several proteins form amyloid fibrils from near-native states, including some lysozyme variants, superoxide dismutase 1, and the polyglutamine-containing protein ataxin-3 (51-54). Structural fluctuations around the protein regions that are involved in intermolecular interactions can lead to the formation of prefibrillar and amyloid aggregates retaining some native-like structure (54-56). Emerging techniques to monitor conformational dynamics, including NMR relaxation dispersion, hydrogen-deuterium (HD) exchange and single molecule fluorescence methods will be important for determining the earliest events leading to the aggregation of globular proteins.

Some proteins are unfolded in the native state, stabilized by high configurational entropy (1). The native state of a number of amyloidogenic proteins, including  $\alpha$ -synuclein, A $\beta$  fragments, and polyglutamine fragments is essentially random coil (57). Mutations or proteolytic cleavage of unfolded fragments can initiate aggregation by altering the conformational fluctuations of the unfolded ensemble, leading to partial folding to metastable monomeric states that are stabilized by intermolecular interactions.

Proteolytic cleavage initiates the aggregation of a number of proteins, including A $\beta$ , Ig light chain and some polyglutamine-containing proteins, and there is evidence that proteolysis is enhanced in some disease states (58-60). Altered conformational dynamics upon mutation may be a precursor for fragmentation, allowing cleavage sites to become more accessible to proteases. Protein fragments are usually more aggregation-prone than the full-length protein variants, indicating that protein regions flanking amyloidogenic sequences can suppress intermolecular interactions.

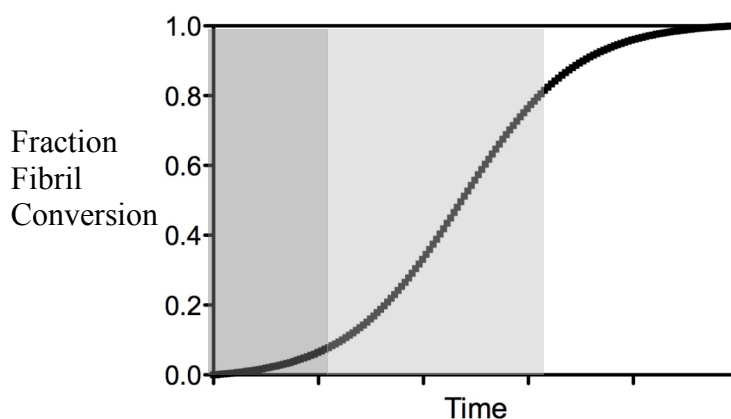
Native state perturbations leading to structural destabilization, local conformational fluctuations and proteolysis can all trigger aggregation cascades. Identifying the initial conformational changes will therefore be important in ultimately suppressing the pathways leading to the formation of amyloid and/or prefibrillar toxic species.

### **1.1.3.2 Kinetics of aggregation**

Amyloid fibrils typically display a nucleated growth polymerization mechanism (61, 62) (Figure 1.3). The rate of aggregation is determined by a nucleation phase, involving a slow transition from the native state to a non-native conformer during the lag time of the reaction (61). Nucleation involves a thermodynamically unfavourable transition to a high-energy conformer; an altered monomeric state or a multimeric nucleus formed by condensation of several subunits (61). Nuclei are unstable and rapidly decay back to the bulk monomeric phase, however, collisions in solution lead to a rapid elongation phase involving monomer addition to growing fibril ends. Dependence on intersubunit interactions for elongation therefore makes the rate of amyloid formation concentration-dependent.

The aggregation kinetics of A $\beta$ , polyQ peptides, and several other amyloid-forming proteins have been described, providing important insights into the aggregation pathways (63-66). Different kinetic parameters obtained for these systems indicate that the species formed during aggregation and their rate of formation depends on the polypeptide sequence and protein context. Altered aggregation kinetics are often observed upon mutation and under variable solution conditions that influence the native state.

Oligomers form in the early stages of aggregation, and they may be on-pathway to amyloid formation, however, are typically thought to form in competing pathways by nucleation-independent mechanisms (62). These species are often observed in the lag phase of the nucleation-dependent amyloid kinetics (Figure 1.3). These species complicate kinetic analyses and therefore the mechanisms of amyloid and prefibrillar oligomer formation are difficult to characterize. In some systems, however, the kinetic scheme can provide important insight into the early molecular events on the pathway to amyloid formation.



**Figure 1.3: Nucleation-dependent aggregation kinetics.** A representative trace of nucleation-dependent aggregation kinetics. The dark grey box shows the lag time in which competing nucleation-independent oligomerisation events may occur, and the lighter grey box highlights the elongation phase.

### 1.1.3.3 Amyloid oligomers and toxicity

Energetic landscapes for amyloid systems are complicated by the population of small aggregates, including oligomeric and protofibrillar species that have been observed by TEM and AFM as spherical or short curvilinear species (67). These are usually unstable, transiently-populated species (Figure 1.1), however, they are of significant interest as they are thought to be central to the cytotoxicity associated with amyloid diseases (4, 49).

The transient nature of oligomeric populations and conformational heterogeneity has limited their high-resolution structural characterization. Limited proteolysis, hydrogen-deuterium exchange and FTIR studies suggest that oligomers are less ordered, have a reduced  $\beta$ -sheet content and enhanced exposure of hydrophobic regions compared to fibrils (68-70). Further spectroscopic techniques, including FRET (70), NMR techniques (71-73), and single-molecule fluorescence (74) also show promise in detecting and characterizing small populations of oligomeric complexes.

In the absence of high-resolution structural data, conformation-specific antibodies have been useful tools to detect oligomeric species (9, 75, 76). The Glabe laboratory has generated a range of monoclonal antibodies with high affinity for fibrillar, prefibrillar, and annular conformations, allowing the detection of these species under different conditions and within cells and tissues (9, 76, 77). These antibodies were raised against A $\beta$  species, however, they also bind oligomers and protofibrils formed by different polypeptides, including  $\alpha$ -synuclein, transthyretin, lysozyme, insulin and prion protein fragments, indicating generic elements of structural organization in oligomers and protofibrils (77, 78). Different classes of oligomers that have variable morphology, size and toxicity have been identified and these may represent alternative association pathways. Fibrillar oligomers have the same structural organization as mature fibrils and may represent species on-pathway to amyloid formation. Prefibrillar oligomers are structurally distinct from amyloid fibrils and are thought to be precursors for curvilinear and annular protofibrils (76, 79).

Some of the biggest unanswered questions in the amyloid field lie in determining which species are toxic to cells and their mechanisms of toxicity. In patients and *in vivo* models, the extent of fibrillar deposition does not strictly correlate with disease (80, 81). There is evidence that A $\beta$ , prion protein, and  $\alpha$ -synuclein

oligomers and/or protofibrils are associated with toxicity (78, 82-85). Similarities in the morphology of protofibrillar and oligomeric species suggest that they may have similar mechanisms of toxicity and it is possible that a number of different species are toxic. The precise mechanisms by which oligomeric species induce toxicity are unknown, however, are thought to include membrane permeabilization, perturbations of redox status, impairment of calcium metabolism and disruption of proteasomal activity, ultimately leading to apoptosis (4, 79, 86).

## **1.2 Polyglutamine Misfolding**

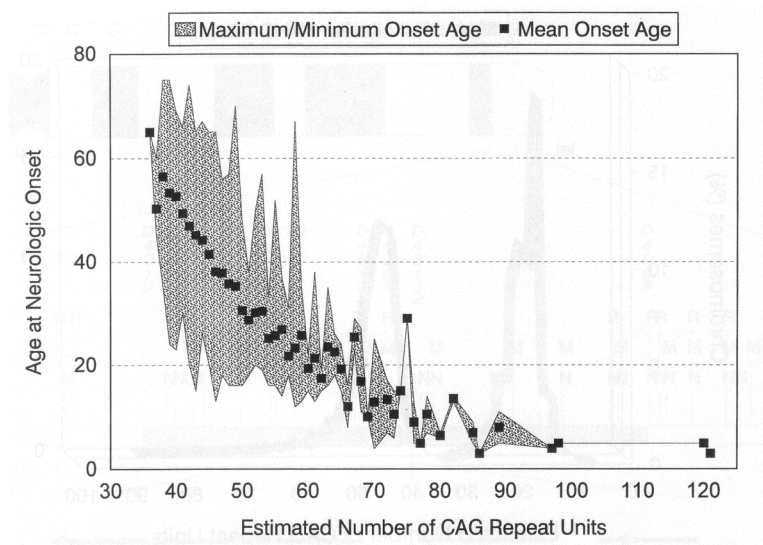
### **1.2.1 Polyglutamine Diseases**

The polyglutamine (polyQ) disorders are a family of nine neurodegenerative diseases and comprise Huntington's disease, spinocerebellar ataxias 1,2,3,6,7 and 17, spinobulbar muscular atrophy (SBMA) and dentatorubropallidoluysian atrophy (DRPLA). PolyQ diseases are progressive and invariably fatal and the nine diseases have variable clinical phenotypes as different neuronal subsets are affected in each disease (87).

The genetic mechanism leading to the polyQ diseases is a trinucleotide repeat expansion of a CAG repeat within an exon of the disease-associated protein, encoding a protein with an expanded polyQ tract. Each disease has a specific pathological threshold (Table 2) above which there is a correlation between repeat length and age-of-disease onset and severity of clinical symptoms (Figure 1.4). CAG repeat expansions are dynamic, showing somatic and germline instability in repeat length (88-90). This leads to intergenerational expansion, resulting in enhanced phenotypes in successive generations.



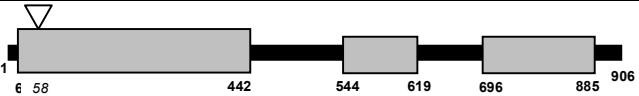
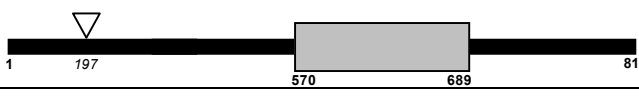
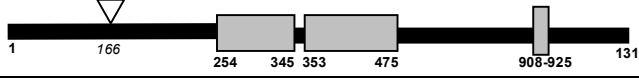
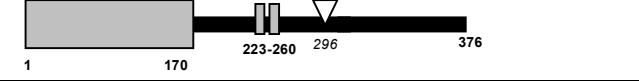
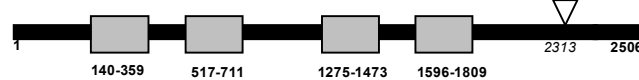
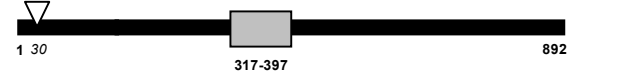
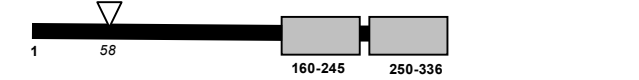


Several hypotheses have been proposed to explain the toxicity associated with expanded length polyQ tracts. It is widely considered that toxic activities of the protein lead to disease. A gain-of-function mechanism was demonstrated with the addition of an expanded polyQ repeat tract to the unrelated hypoxanthine-guanine phosphoribosyltransferase (HPRT) inducing protein aggregation and cell death in a mouse model (92). The implications of this study were further substantiated by *in vitro* findings showing similar gain-of-function aggregation mechanisms (93, 94). Recent data also suggest that RNA-derived toxicity may be a further pathway involved in disease (95). The remainder of this chapter will focus on the protein misfolding and aggregation pathways involved in polyQ pathogenesis.



**Figure 1.4: The relationship between the age of disease onset and the polyQ repeat length in patients with Huntington's disease.** There is an inverse exponential correlation between repeat length and age at onset. (Figure from ref (91)).

**Table 2: Domain architectures of the nine polyQ disease proteins**

| Disease    | Protein           | PolyQ disease threshold | Protein architecture*  | PolyQ flanking sequences (N'-C')      | Refs       |
|------------|-------------------|-------------------------|--|---------------------------------------|------------|
| DRPLA      | Atrophin-1        | >49                     |    | PPVSTHHHHH(Q) <sub>n</sub> HHGNSGPPPP | (96, 97)   |
| HD         | Huntingtin        | >36                     |    | MKAFESLKSF(Q) <sub>n</sub> PPPPPPPPP  | (98, 99)   |
| SBMA       | Androgen Receptor | >40                     |    | APPGASLLLL(Q) <sub>n</sub> ETSPRQQQQQ | (100, 101) |
| SCA1       | Ataxin-1          | >41                     |    | LSQTPGHKAE(Q) <sub>n</sub> HLSRAPGLIT | (102, 103) |
| SCA2       | Ataxin-2          | >35                     |    | YGPLTMSLKP(Q) <sub>n</sub> PPPAANVRK  | (104, 105) |
| SCA3 (MJD) | Ataxin-3          | >45                     |    | RKRREAYFEK(Q) <sub>n</sub> GDLSGQSSHP | (106, 107) |
| SCA6       | CACNA1A           | >21                     |   | IRKAGGSGPP(Q) <sub>n</sub> AVARPGRAAT | (108, 109) |
| SCA7       | Ataxin-7          | >37                     |  | AGGAAAAAA(Q) <sub>n</sub> PPPPQPQRQQ  | (110, 111) |
| SCA17      | TBP               | >47                     |  | SLSILEEQQR(Q) <sub>n</sub> AVAAAAVQQS | (112, 113) |

Abbreviations: DRPLA- dentatorubropallidolulysian atrophy; HD- Huntington disease; SBMA- Spinobulbar muscular atrophy; SCA- spinocerebellar ataxia; MJD- Machado Joseph disease; CACNA1A- alpha1 voltage-dependent calcium channel; TBP- TATA-box binding protein. \*Protein architectures and domain boundaries were identified using Pfam. Triangles indicate the position of the polyQ (italicized numbers indicate the position of the first glutamine in the repeat tract). Figures are not to scale. Black regions indicate structurally undefined regions (the domain architecture of atrophin-1 has not been described) and grey boxes indicate defined domains.

---

**1.2.2 The Physical Behaviour and Aggregation of PolyQ**

---

The large size of the disease proteins (Table 1), as well as their relative insolubility have led to difficulties in the recombinant production and isolation of most of the pathological proteins (114), and as a result, much of the current knowledge of polyQ-dependent aggregation and toxicity has been derived from peptides and model proteins.

The common pathological features of the polyQ diseases support the hypothesis that these diseases share a common mechanism of toxicity entirely dependent on the presence of an expanded polyQ region. The sidechain homogeneity of the consecutive polar glutamines lead Max Perutz to propose the polar zipper model for polyQ segments (115). The polar zipper model posits that  $\beta$ -strands form intermolecular hydrogen bonds between their main chain and side chain amide groups forming  $\beta$ -pleated sheets that are thermodynamically stable (115). This model is supported by experimental evidence showing a  $\beta$ -sheet X-ray diffraction pattern for polyQ peptides and other evidence indicating that aggregates are  $\beta$ -sheet rich (115).

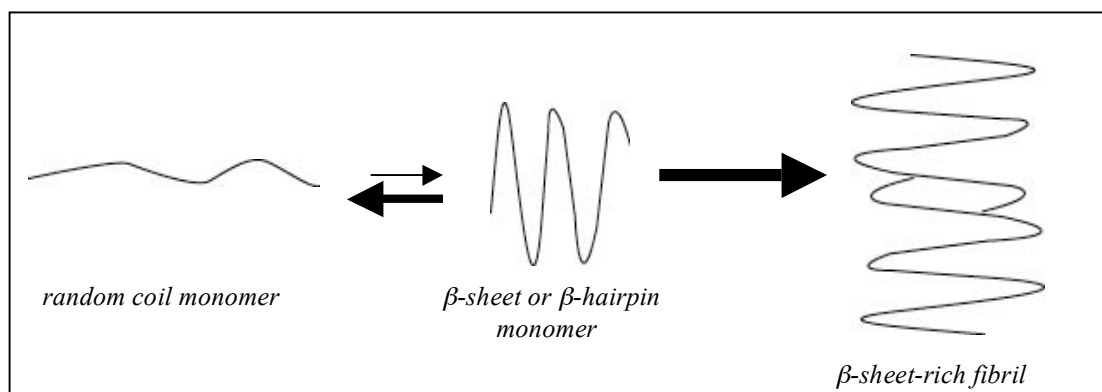
PolyQ homopeptides form amyloid-like aggregates, and homopeptide studies have been exceptionally useful in revealing some of the fundamental properties of polyQ aggregation. Wetzel and colleagues established an inverse correlation between polyQ repeat length and the rate of peptide aggregation (63). This is similar to the relationship between repeat length and age-of-onset in Huntington's and other polyQ diseases, however, the threshold to aggregation of the peptides is significantly less than the threshold to disease onset. Even a Q15 peptide was shown to form amyloid-like aggregates with a similar straight, unbranched fibrillar morphology to aggregates of longer polyQ fragments (116). When situated within the context of a full-length protein, the repeat length threshold to aggregation more closely resembles that observed in

disease (94). These data suggest that the position of the polyQ tract within a protein and the physical nature of the surrounding sequences modulate the aggregation reaction.

PolyQ homopeptides aggregate with nucleation-dependent kinetics, similar to many other amyloid-forming proteins (61). Nucleation involves a localized conformational change within the monomeric polyQ sequence (63). This kinetic mechanism for polyQ aggregation has also been demonstrated in the host protein environment with two further reports describing monomeric nuclei as the trigger for aggregation (66, 93). These data suggest that the biophysical behavior of the monomeric polyQ-containing proteins is fundamental to aggregation. In agreement with this it has been demonstrated that targeted stabilization of the polyQ tract by the peptide QBP1, which binds to polyQ with high affinity, leads to inhibition of aggregation (117, 118).

The conformation of the monomeric polyQ nucleus has not been experimentally described. It has been predicted that the critical concentration of misfolded polyQ species required to initiate aggregation is in the nanomolar range (116), and the favorability of elongation post-nucleation is extremely rapid. As such, the lifetime of the misfolded conformer under native conditions would be too short to be detectable by ensemble experimental techniques, such as NMR and CD. It is reasonable to assume that only single-molecule techniques have the sensitivity required to monitor such small populations of conformations. Considering the  $\beta$ -sheet architecture of fibrils, a plausible hypothesis is that monomeric polyQ sequences misfold to a  $\beta$ -sheet or  $\beta$ -hairpin conformation (Figure 1.5). Molecular simulations investigating the conformational ensembles of monomeric polyQ have obtained varied results, which could be attributed to the variable experimental conditions utilized (119-121). Proposed models include  $\alpha$ -sheet,  $\beta$ -helix, and compact random coil conformations (119-121). The variability between these studies suggests that the conformation of the polyQ tract is sensitive to

local environment and solution conditions, and the rapid conformational fluctuations of the polyQ tract could lead to the sampling of varied conformations under different conditions.



**Figure 1.5: Model for the misfolding of polyQ homopeptides.** Native polyQ exists in a random coil conformation, undergoing an unfavorable conformational change to a proposed  $\beta$ -sheet or  $\beta$ -hairpin monomer that acts as an aggregation nucleus. Rapid fibril elongation then occurs.

A number of experimental investigations suggest that the predominant conformation of monomeric polyQ is a random coil conformation, and importantly, that this conformation is independent of repeat-length, ie, pathological and non-pathological repeat tracts alike are both random coil in conformation (94, 122). Structural investigations demonstrated that the polyQ-binding antibody 1C2 binds pathogenic and non-pathogenic lengths with similar affinity, however, longer repeats harbor an increased number of binding sites, consistent with a model of a repeating uniform conformation despite repeat length (123, 124). Considering interglutamine contacts are unfavored in the random coil conformation, the thermodynamic favorability of monomeric polyQ is derived from high configurational entropy of the disordered chain (125). The polar nature of the glutamine sidechains mean that in addition to water, side chain moieties can also solvate the backbone, further increasing the possible number of

backbone configurations and stabilizing the random coil conformation (125). The unfavorable equilibrium between the random coil and  $\beta$ -sheet rich monomeric conformation could be reasoned thermodynamically by a minimal enthalpic gain in comparison with the large entropic penalty of folding to a structured conformation. The increased chain length of longer polyQ segments would contribute a higher entropic term for longer polyQ; counterintuitive considering the accelerated aggregation kinetics of pathological length polyQ. This effect is likely offset by the increased thermodynamic contribution associated with the stacking of longer polyQ tracts within the fibril. There are several other examples of amyloid-forming proteins that prefer an unfolded/random coil monomeric conformation (126, 127), and a number of studies demonstrate that proteins and peptides that have a preference for random coil are capable of aggregating readily (1).

### **1.2.3 The Protein Context of PolyQ**

The nine polyQ disease proteins vary widely in terms of their sequence length, domain architecture (Table 1), proposed function, cellular localization and patterns of neuronal expression. Presumably, these protein-specific differences give rise to the particular hallmarks of each disease (87). This raises the question: are the specific sequence contexts of polyQ linked to their aggregation behavior? There are several proposed mechanisms by which the flanking regions modulate polyQ aggregation (Figure 1.6).

The role of the flanking sequences on polyQ aggregation behavior has been investigated by Nozaki *et al* (128). This study investigated the aggregation of fusion proteins comprising a pathological length polyQ and the flanking eight N-terminal and nine C-terminal amino acids from ataxin-2, Huntingtin (Htt), atrophin-1 and ataxin-3 (128). Consistent with HD and SCA2 having slightly lower polyQ repeat thresholds

compared to DRPLA and SCA3 is the finding that the Htt and ataxin-2 variants had higher aggregation propensity compared to the other two proteins, indicating that the specific properties of these flanking sequences can modulate aggregation. Htt has a proline-rich region directly C-terminal to the polyQ tract which has been shown to suppress aggregation by a mechanism in which the conformation of the polyQ is altered (129). The presence of the polyproline (polyP) motif leads to a coincident reduction in toxicity in a yeast cell model (130). The suppressive effect could be contributed by the rigidity of the polyP as proline-rich regions are intrinsically capable of forming P11-helices and this may impact on the conformation of adjacent regions. Owing to structural constraints prolines do not readily adopt  $\beta$ -structures, and prolines have been conserved in a number of contexts for their ability to act as a breaker of  $\beta$ -sheet and amyloid formation (131). Considering the suppressive effect of polyP on aggregation, it is possible that these sequences have been evolutionarily conserved to have a gatekeeper role in preventing aggregation. A recent study by Wetzel and colleagues suggests that the region N-terminal to the polyQ tract in Htt enhances aggregation by a mechanism that is dominant over the suppressive effect of the polyP (132). It would be interesting to determine whether the flanking sequences of any of the other disease proteins have similarly competing effects on aggregation.

Some of the other disease proteins have adjacent regions that are rich in single amino acids. Ataxin-7 also has a proline-rich region, TBP has an alanine-rich region C-terminal to its polyQ tract, and atrophin-1 has a penta-histidine sequence N-terminal to the polyQ (Table 1). Aside from polyP, the relationships between these homogeneous flanking regions and polyQ are yet to be determined. Sequentially homogeneous flanking regions may be conserved in these regions due to their ability to modulate aggregation by affecting conformation, solubility or other factors.

Several studies have found a link between protein solubility and aggregation (128, 133-135). The charge of flanking amino acid sequences affects aggregation (128), and several groups have utilized solubility-enhancing charged flanking residues to slow the aggregation of polyQ peptides (115, 135). A number of groups have also utilized fusion tags in order to express and purify recombinant polyQ-containing proteins. A GST-Htt exon1 fusion protein with a pathological length polyQ repeat is aggregation-incompetent prior to cleavage of the soluble GST tag (122, 136). The GST tag has also been shown to reduce toxicity in a yeast model (130). Similarly, an ataxin-3 variant comprising a pathological Q78 repeat could only be purified when fused to a soluble fusion tag to avoid aggregation (137). Combined, these data suggest that there is a link between the total electrostatic charge of the protein and the propensity for polyQ aggregation. This is consistent with the wider protein aggregation field, as protein solubility and gene expression levels are linked to a protein's aggregation propensity (138). A study of regions of high amyloid propensity in the human proteome indicated that the regions flanking the amyloid-prone sequences are enriched in prolines and charged residues, suggesting that enhancing the local solubility may be a general evolutionary mechanism to avoid aggregation (139).

#### **1.2.4 PolyQ Expansion and Protein Stability**

At the structural level, most polyQ tracts are located within unstructured/topologically undefined regions N- or C-terminal to a structured domain (140). The polyQ tract in at least seven of the nine disease proteins is located in a flexible solvent-exposed region of the protein (Table 1). This raises the question: is polyQ tract location significant?

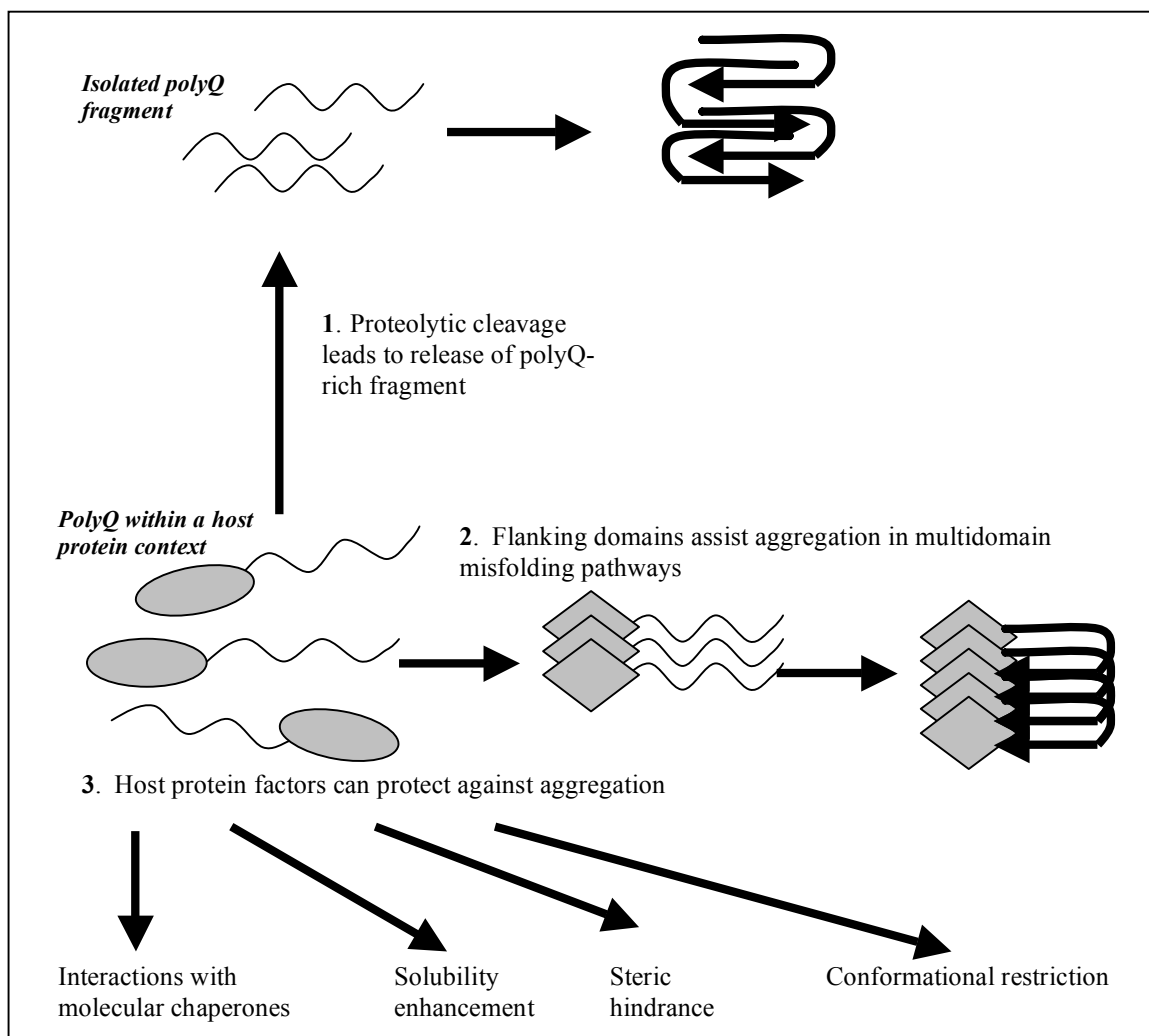
One hypothesis is that the dynamic fluctuations of the polyQ region affect the local conformation, altering the stability and/or conformational dynamics and this has



been tested in a number of systems with varied outcomes (51, 93, 94, 132, 141). It was recently demonstrated that polyQ length alters the conformation of the adjacent 17 amino acid N-terminal domain of Htt exon1, which exists in a more extended conformation in the presence of a pathological length repeat (132). This is consistent with molecular dynamics simulations which suggest that this domain can exist in two stable conformations (98). These studies raise the possibility that the equilibrium between these conformations is dependent on the polyQ repeat length. Ignatova and Gierasch showed that a long polyQ tract can destabilize more distantly located domains, showing that the cellular retinoic acid binding protein 1 (CRABP1) domain is perturbed when fused to the N-terminus of Htt exon1 (93). Similarly, polyQ expansion of the androgen receptor (AR) is accompanied by secondary structural changes in the N-terminal flanking domain (142). Two studies from our group indicate that structural destabilization is not always repeat-length dependent. Pathological and non-pathological polyQ lengths have an equivalent effect on the thermodynamic stability of ataxin-3 (51, 143). The work of this doctoral thesis shows that the structural integrity of a juxtaposed  $\alpha$ -helical domain is unchanged in the presence of a pathological length polyQ tract (94). These varied results suggest that structural perturbations may contribute to pathogenesis in some diseases, but are not essential for aggregation and may not be a necessary disease component.

The significance of repeat tract location is highlighted by a number of studies in which ectopic polyQ have been placed within loops and/or linking regions of structured domains (94, 141, 144). Insertion of short polyQ tracts within the chymotrypsin inhibitor 2(CI2), *Staphylococcus aureus* Protein A(SpA) and Myoglobin(Mb) domains caused significant structural destabilization. In the CI2 model, the decrease in thermodynamic stability is not repeat length dependent (144), whilst there was a

stepwise increase in destabilization in the SpA and Mb proteins with increasing polyQ length (94, 141). Compared to terminally-located polyQ, in these cases polyQ expansion may induce an additional thermodynamic penalty if protein regions located N- and C-terminal to the polyQ form stable interactions in the native state. Most of the disease polyQ tracts are not located within a structured domain, obscuring the interpretability and relevance of these models to the pathological scenario.



**Figure 1.6. Proposed pathways by which the host protein can modulate the aggregation of polyQ proteins.** 1) Proteolytic cleavage leading to release of a polyQ-rich fragment from flanking protein domains can enhance aggregation. 2) Additional aggregation-prone regions of the host protein may lead to multidomain aggregation mechanisms. 3) Properties of the host protein may help to avoid aggregation. For example, chaperones and/or other binding partners may interfere with the aggregation mechanism. Host protein regions may also protect against aggregation by enhancing total protein solubility, sterically hindering polyQ intermolecular interactions, and/or restricting the polyQ conformational changes required for fibril nucleation and/or elongation.

---

**1.2.5 Proteolytic Cleavage: A Protective role of Protein Context**

---

Consistent with data indicating protein solubility as a factor that influences aggregation is evidence that a number of the disease proteins require proteolytic cleavage of the polyQ-containing fragment from the host protein prior to aggregation (145). Analysis of patient samples and animal models showed that neuronal inclusions, in some cases, comprise not only the full-length polyQ-containing proteins, but also proteolytically-cleaved fragments of these proteins (146, 147). Full-length Htt comprises over 3000 residues, and the poly(Q) tract is located 17 amino acids from the N-terminus of the protein (99). Nuclear inclusions of HD patients consist largely of the N-terminal Htt exon1 fragment, which comprises the poly(Q) repeat and 65 extraneous residues (148). Cleaved fragments of some other disease proteins, including atrophin-1, androgen receptor, and ataxin-7 have also been identified, suggesting that proteolytic cleavage is also involved in the pathogenic cascades of these proteins (145, 149). The work described above raises the question: are proteins with a pathological length polyQ repeat preferentially cleaved? Due to the context-dependent effect of a pathological length polyQ tract on the structure and stability of flanking regions, this cannot be ruled out as a disease mechanism. A number of studies have investigated the kinetics of cleavage of the disease proteins. Htt with an expanded polyQ tract is cleaved at a faster rate by apopain (150), however, polyQ repeat length did not alter the cleavage efficiency of Htt, atrophin-1, ataxin-3 or AR by caspase-3 (145). This variability may arise from the altered specificity of different proteases and/or variable effects of polyQ on the proteolytic susceptibility regions.

Proteolytic cleavage precedes the aggregation of polyQ fragments expressed and purified in the context of a fusion protein (122, 136), suggesting that a large host protein context is inhibitory to aggregation. Considering the protein in isolation, this raises the

possibility that total protein solubility and/or steric hindrance due to the presence of neighbouring domains can modulate the formation of polyQ intermolecular interactions.

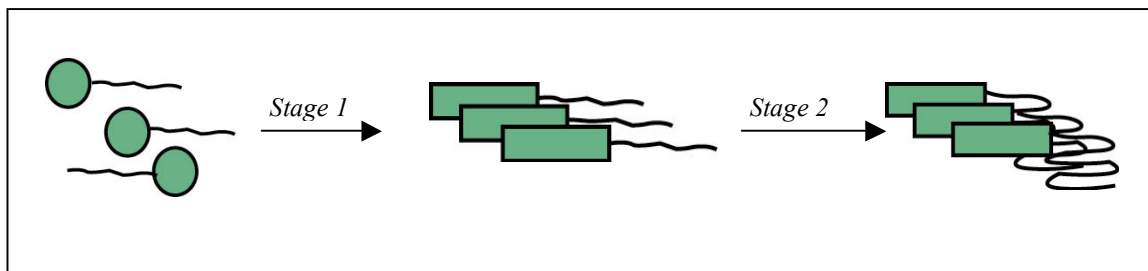
Evidence that the properties of polyQ flanking residues can suppress polyQ-dependent aggregation indicates a protective role of the host protein in some cases. Exploiting and/or mimicking the effects induced by the host protein in these cases could present potential therapeutic options.

### **1.2.6 Multidomain Aggregation: Host Protein Domains that Assist PolyQ Aggregation**

Contrary to the idea that regions of the host protein can protect against polyQ aggregation comes recent evidence for a number of the disease related polyQ proteins that regions flanking the polyQ tract can in fact assist and/or enhance the rate of polyQ-dependent fibril formation and can influence the type of aggregate formed (132, 151).

A two-stage aggregation pathway was described for ataxin-3 (151). In this model, the first stage is mediated by intermolecular interactions of the globular N-terminal Josephin domain, with subsequent polyQ interactions leading to the formation of SDS-insoluble, fibrillar aggregates (Figure 1.7) (151). Protein variants harboring a non-pathological length polyQ and the isolated Josephin domain also aggregate *in vitro*, forming small curvilinear aggregates that can be solubilised by SDS (151, 152). Intracellular aggregates of isolated Josephin have not been observed, although there is evidence that non-pathological ataxin-3 forms aggregates in mammalian cells, suggesting that the Josephin stage of ataxin-3 aggregation is relevant *in vivo* (153).

A recent study suggests that a similar mechanism is involved in the aggregation of Htt exon1 peptides (132). The N-terminal 17 amino acids were shown to interact and initiate the formation of oligomeric species with the aggregate core composed of the N-terminal domain, surrounded by exposed polyQ tails (132). Description of these species



**Figure 1.7: Model of the two-stage aggregation mechanism of ataxin-3.** The green circles represent the Josephin domain. Stage 1 is rapid and involves conformational changes leading to Josephin self-association. Ataxin-3 with an expanded polyQ tract can then form stable SDS-insoluble fibrils mediated by polyQ intermolecular interactions in the second stage of aggregation.

is of interest considering that the presence of the N-terminal domain also modulates toxicity (132, 154) which can be suppressed by the addition of antibodies specifically targeting the N-terminal 17 amino acids (155). When fused to a pathological length polyQ, the N-terminal domain is in a more extended aggregation-prone conformation, suggesting a perturbing influence of the expanded polyQ (132) (see earlier section). Aggregation of the N-terminal domain is a rate-determining step for subsequent polyQ interactions; polyQ peptides fused to the N-terminal 17 amino acids aggregate much more rapidly than isolated polyQ peptides.

A study of a CRABP1-Htt exon1 fusion protein suggests that the perturbing influence of polyQ expansion in the Htt exon1 context can be transmitted beyond the 17aa N-terminal domain, showing that the  $\beta$ -barrel architecture of CRABP1 is disrupted when fused to the N-terminus of Htt exon1 with a polyQ expansion (93). The morphology of the CRABP1-Htt exon1 aggregates evolves over time, early timepoints are SDS-soluble, however this changes as the fibril core is increasingly dominated by stable polyQ interactions (156). This protein contains at least three potential aggregating regions; CRABP1, the N-terminal Htt domain and the polyQ. It would be interesting to

delineate the aggregation mechanism further to determine whether the conformational changes preceding the intermolecular interactions of the Htt N-terminal domain can still occur in the presence of CRABP1 and/or which regions of the fusion construct dominate various stages of the evolving fibril core.

There is also considerable evidence to suggest that ataxin-1 aggregates by a similar multidomain mechanism (157, 158). The C-terminal AXH domain of ataxin-1 is a highly dynamic, chameleon-like domain with an environmentally-sensitive structure that can spontaneously aggregate independent of the polyQ tract, both *in vitro* and *in vivo* (157, 158). The AXH domain has a functional role, participating in protein-protein interactions leading to transcriptional repression, and it has been demonstrated that these interactions are altered in the context of a pathological length polyQ (159-162), suggesting that the polyQ tract may have a further effect on the fold and/or conformational fluctuations of the AXH domain. Similar to Josephin in ataxin-3 and the N-terminal domain of Htt exon1, the AXH domain has a rate-modulatory role on polyQ aggregation, as deletion or replacement of the AXH domain retards the polyQ aggregation stage (163). Complete characterization of the *in vitro* aggregation kinetics has not yet been performed for ataxin-1 as protein size limitations have hindered the isolation of recombinant variants, however, evidence thus far indicates that a multistage mechanism is likely to occur.

The loss of native structural integrity for sequences with a high propensity for  $\beta$ -sheet formation commonly leads to amyloid formation (164). These examples indicate that protein regions outside of the polyQ tract may be sensitive to perturbations induced by the conformational fluctuations of the polyQ region. As a result, these vulnerable regions can provide additional pathogenic aggregation-prone elements. The  $\beta$ -barrel architecture of CRABP1 combined with the structural destabilization induced by the

pathological length polyQ are likely to make this a susceptible domain to participate in a multidomain aggregation pathway. The Josephin domain of ataxin-3 and the AXH domain of ataxin-1 both have considerable  $\beta$ -sheet components (165, 166), and the altered stability and/or dynamics of these domains upon polyQ expansion could lead to enhanced  $\beta$ -sheet exposure. Regions of random coil also have a high aggregation propensity (1); therefore, the proposed random coil structure of the Htt N-terminal domain (132) may enhance the susceptibility of this region to polyQ-induced perturbation leading to multidomain aggregation.

Initial aggregation of non-polyQ regions is not a necessary step that precedes polyQ intermolecular interactions; polyQ homopeptides spontaneously aggregate *in vitro* (63) and a polyQ-containing model protein aggregates by a simple mechanism involving only the polyQ tract (94). Multidomain pathways, however, seem the most biologically relevant mechanisms for polyQ aggregation: is it a coincidence that the aggregation pathways for all of the disease proteins analyzed thus far have additional non-polyQ aggregation steps? Is this a feature that enhances aggregation and/or disease susceptibility, and do these alternative interactions lead to pathways that are important in toxicity? Small protofibrillar and oligomeric aggregates have been observed when the polyQ tract is located within ataxin-3 and Htt exon 1 (132, 152, 167), and it would be of interest to determine whether these species have a role in toxicity. These are some of the questions that will only be answered by investigating the aggregation pathways of the remaining polyQ disease proteins; a very difficult task given the problems associated with obtaining recombinant variants of large proteins. Clearly defining the intracellular and physiological relevance of alternative aggregation stages and determining whether the early structures that result from these steps are involved in toxicity would provide a

good basis for rational therapeutic targeting of aggregation-susceptible non-polyQ protein regions.

Multidomain aggregation mechanisms do not appear to be unique to the polyQ protein family. There is evidence that PABPN1, the protein implicated in OPMD, has more than one aggregation-prone sequence element. The OPMD mutation is the expansion of a polyalanine (polyA) sequence that leads to fibrillar aggregation. Truncated variants of PABPN1 lacking the polyA domain also aggregate, indicating that there are secondary regions prone to aggregation (168). The yeast prion Ure2p also has multiple aggregation-prone regions; the N-terminal glutamine/asparagine rich prion domain, and a region toward the C-terminus of the protein have been shown to aggregate independently (169). Given the homopolymeric nature of the polyalanine in PABPN1 and the glutamine/asparagine-rich topology of the N-terminal Ure2p domain, it is not unreasonable to speculate that the aggregation mechanisms for these proteins may be closely aligned with that of the polyQ proteins.

### **1.2.7 Modulation of Misfolding, Aggregation and Toxicity through Interaction Networks**

At the cellular level polyQ protein aggregation leads to activation of the protein quality control machinery - molecular chaperones and the ubiquitin proteasome system- indicating that misfolding is central to disease. These cellular protective mechanisms attempt to maintain homeostatic balance in response to stress induced by protein misfolding. Nuclear inclusions formed in polyQ and other amyloid diseases are often ubiquitinated and recruit molecular chaperones and other components of the quality control machinery, indicating involvement of these pathways in pathogenesis. However, there are several lines of evidence that isolated polyQ is an inefficient substrate for these systems, and protein context-specific differences in recognition and interactions



suggest that the host protein affects the efficiency of chaperone and proteasomal subunit recognition.

Molecular chaperones have a critical role in recognizing and refolding misfolded proteins. Three main categories of chaperones; the Hsp70 class, the chaperonins and the small heat shock proteins (sHsps) have been shown to modulate polyQ-derived aggregation and toxicity.

Chaperones can recognize and prevent the toxicity of some polyQ-containing proteins and data indicate that chaperone activity is context-dependent (170, 171). For example, the sHsp  $\alpha$ B-crystallin suppresses ataxin-3-induced toxicity, however, does not alter the phenotype caused by Htt overexpression in a *Drosophila* model (170, 171). Considering that the uniting feature of the Htt and ataxin-3 systems is the polyQ region, these data also suggest that homogeneous polyQ is a poor substrate for molecular chaperones. Further supporting this is the observation that Hsp70 had no effect on polyQ-induced toxicity in the absence of its co-chaperone Hsp40 (172, 173).

There is evidence that the Hsp70/Hsp40 chaperone complex inhibits the formation of spherical and annular oligomeric structures formed by Htt exon1, enhancing favourability for monomer (174). The chaperonin TRiC appears to suppress Htt aggregation by a similar mechanism (175, 176)(177). Considering the proposed multidomain aggregation mechanism, and the hypothesis that interactions leading to oligomer formation occur largely in the N-terminal domain of Htt, these data raise the possibility that these chaperones inhibit the early stages of Htt aggregation that involve the N-terminal domain (132, 176, 177).

Several genetic screens have identified modifiers of different polyQ disease phenotypes, with a number of chaperones having strong suppressive effects (170, 178, 179). A comparative *Drosophila* study of SCA1 and HD modifiers identified two heat-

shock response proteins that suppress the SCA1 phenotype, however, have no effect on the HD model (179). A genetic screen of modifiers of the ataxin-3 phenotype in *Drosophila* found that a sHsp has an enhanced repressive effect on aggregation of the full-length protein compared to a C-terminal fragment (170). The variable conformational relationships between the polyQ and flanking domains, and the partial unfolding of flanking domains leading to the multidomain aggregation pathways may be responsible for the context-dependent nature of chaperone recognition of polyQ-containing substrates.

Preferred substrates of molecular chaperones have exposed hydrophobic residues. Proteins have evolved such that hydrophobic residues are buried in their native conformation, and these can become exposed under destabilizing conditions with high potential to become aggregation-prone. The hydrophilic character of glutamines could explain the poor affinity of chaperones for pure polyQ substrates. However, considering the perturbing influence of the polyQ tract on some host proteins and resulting multidomain aggregation pathways, specific sequence elements outside of the polyQ region would be likely targets for chaperone binding. As molecular chaperones are present within mammalian cells and expression is enhanced in disease, one hypothesis is that the misfolding of the host protein in these cases is in fact protective; providing an exposed hydrophobic patch that can be recognized by chaperones. Further *in vivo* studies of the aggregation pathways and coincident chaperone overexpression may help to address the cellular relevance of multidomain misfolding.

Chaperone activation leads to pathways whereby misfolded proteins can be shielded, disaggregated and/or refolded. However, there is no polyQ consensus fold, and it is unreasonable that the polyQ segment would be refolded to a more stable native conformation. The cellular degradative machinery therefore offers an alternative

pathway for the elimination of aberrant protein conformers. The ubiquitin proteasome system normally functions to remove excess damaged proteins from the cytosol and nuclei leading to proteolytic degradation and recycling of constituent amino acids (180, 181). This system, however, is thought to have limited efficiency in digesting polyQ proteins (182-184). A number of studies indicate that proteasomes partially digest mutant polyQ-containing proteins (183-185). Isolated polyQ sequences, however, are indigestible by mammalian proteasomes (182) and there is evidence suggesting that undigested polyQ further disrupt cellular homeostasis by blocking proteasomal activity (183, 186). Flanking residues can be cleaved from polyQ fragments (182) and considering that the properties of the flanking sequence may retard polyQ aggregation, and that polyQ fragments are hypothesized to have enhanced toxicity compared to some full-length proteins, this may promote toxicity.

There is evidence that polyQ proteins form complexes with proteasomal subunits with variable affinity. For example, proteasomal components complex stably with Htt exon 1 aggregates, yet interact only transiently with ataxin-1 aggregates (187, 188).

The protein context not only dictates associations of polyQ-containing proteins with cellular quality control components, but with many other cellular proteins. A number of the disease proteins have known or proposed roles in transcriptional regulation (160, 189-191). It is thought that transcriptional dysregulation occurs by recruitment of transcription factors and other proteins to intranuclear inclusions.

Interactions of the host protein with cellular binding partners may or may not be affected by repeat expansion (192-194), consistent with the hypothesis that the host protein can be perturbed by an expanded polyQ tract. Several studies indicate that regions flanking the polyQ tract determine the protein interaction networks. Genetic

screens have identified variable modifiers for SCA1, HD and SCA3 disease models, suggesting that the proteins have different affinity for cellular binding partners (170, 179). Overexpression and/or knockdown of many of the interacting genes can alter inclusion formation and cell survival, highlighting protein interaction networks as a critical modulator of toxicity (170, 179). There is evidence that full-length and truncated variants of the same protein, ataxin-3, specifically interact with different intranuclear proteins (170, 185), which leads to variations in the toxicity induced by the ataxin-3 variants. As the ataxin-3 fragments comprise the polyQ region and immediately flanking protein regions, the variable interactions and toxicity observed in this system highlight the modulatory role of the entire host protein, including regions distantly located from the polyQ.

### **1.3 Aims of this Project**

At the commencement of this study, independent investigations into the role of flanking domains in polyQ aggregation had been limited. The *in vitro* characterization of the polyQ disease proteins has been limited by the difficulty in obtaining purified recombinant variants of these proteins. Some previous studies have utilized polyQ-containing model proteins to determine the effect of a polyQ tract on the stability of adjacent protein regions. The polyQ tract in most of the disease proteins is located in a flexible protein region N- or C-terminal to a domain, and previous models have not considered this. Also, there is no description of the atomic-level structural changes caused by a pathological length polyQ to adjacent domains. There is considerable evidence that intrinsic properties of host proteins, and interactions with molecular chaperones can modulate polyQ-dependent aggregation. However, no systematic *in vitro* studies have characterized this. Determining the role of flanking protein regions in

---

misfolding and aggregation will therefore be valuable in understanding the molecular pathways leading to the polyQ diseases.

The aims of this project are:

- 1.* To construct a model polyQ protein system, necessitating the development of a novel molecular cloning system for amplification of polyQ-encoding sequences.
- 2.* To investigate the aggregation, stability, and structure of the polyQ model proteins.
- 3.* To construct model polyQ protein variants with varying domain architectures in order to investigate the influence of repeat tract location and domain architecture on polyQ-dependent aggregation. Further to this, a bioinformatic analysis of the domain architectures of all human polyQ proteins will be performed.
- 4.* To investigate the effect of a molecular chaperone on polyQ aggregation.

---

## **Chapter 2**

### **Materials and Methods**

## **2.1 Reagents**

Enzymes used for molecular biology were obtained from Promega and New England Biolabs. Phenylmethyl sulphonyl fluoride (PMSF), lysozyme, Thioflavin T (ThT),  $\beta$ -mercaptoethanol (BME), ethylenediaminetetraacetic acid (EDTA), guanidium hydrochloride (GuHCl), guanidinium thiocyanate (GuSCN) and ampicillin (Amp) were all purchased from Sigma Aldrich Co. The ThT, GuHCl and GuSCN were prepared analytically by weight and passed through a 0.22  $\mu$ m filter. The concentration of GuHCl was determined from the refractive index of the prepared solution (195). Benzenesulfonyl fluoride hydrochloride (AEBSF) under the trade name Pefabloc and deoxyribonuclease I (DNase I) were purchased from Roche. The Isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) was prepared analytically by weight and purchased from Astral scientific. The N,N,N',N'-Tetramethylethylenediamine (TEMED) and ammonium persulfate (APS) were obtained from ICN. Enzymatic chemiluminescence reagents (ECL) were obtained from Amersham.

## **2.2 Microbiological techniques**

### **2.2.1 Growth Media**

The media used for bacterial growth was 2YT [16 g/L tryptone, 10 g/L yeast, 5 g/L NaCl], and also on occasions, SOC media [20 g/L tryptone, 5 g/L yeast, 0.5 g/L NaCl, 2.5 mM KCl, 20 mM glucose (added after autoclaving)] was used.

Agar plates were prepared by the addition of 1.5 % (w/v) agar to 2YT medium prior to autoclaving. Amp was added to the media at a concentration of 0.1 mg/mL once the solution had cooled to below 50 °C post-sterilization.



### **2.2.2 Bacterial Strains**

The bacterial strain JM107 was used for general purpose cloning and plasmid maintenance. NovaBlue Competent Cells (Novagen) were used for transformation of products of Quikchange site-directed mutagenesis. For protein expression, BL21(DE3) was used for the pET21a (SpA-polyQ variants) and pLIC vectors (196) (ataxin-3 variants).

### **2.2.3 Oligonucleotide Design**

All oligonucleotides were purchased from Geneworks (Australia). 40 nmol of freeze-dried PCR/Sequencing grade oligonucleotides were obtained and reconstituted in sterile ddH<sub>2</sub>O to a final concentration of 100 µM and stored at -20 °C.

#### **Oligonucleotides**

##### **SpA-cpolyQ amplification**

|                             |           |                                   | <u>Melting<br/>Temp.</u> |
|-----------------------------|-----------|-----------------------------------|--------------------------|
| T7 prom 5'                  | (5' – 3') | cccgcgaaattaatacactcactatag       | 58 °C                    |
| SpAcQ <sub>n</sub> Bsg1 3'  | (5' – 3') | atgcgtgcaggatcctcaccggggtgctgc    | 71 °C                    |
| SpAcQ <sub>n</sub> BseR1 5' | (5' – 3') | tgatgcgcaggcgaggagaccgcggcagca    | 71 °C                    |
| SpAcQ <sub>n</sub> BseR1 3' | (5' – 3') | tgtctgccgcgggtctcctcgctgcgcacatca | 71 °C                    |

##### **SpA-ipolyQ amplification**

|                             |           |                                    |       |
|-----------------------------|-----------|------------------------------------|-------|
| SpAiQ5ins 5'                | (5' – 3') | agcctgaaagatgatccgcagcagcagcagca   | 76 °C |
|                             |           | gagccagagcgcgaatctg                |       |
| SpAiQ5ins 3'                | (5' – 3') | cagattcgcgctctggtctgctgctgctgctgc  | 76 °C |
|                             |           | ggatcatctttcaggct                  |       |
| T7 prom 5'                  | (5' – 3') | As above                           |       |
| SpAiQ <sub>n</sub> Bsg1 3'  | (5' – 3') | gcttcgccagcagagtgcagttcgcgctctgg   | 73 °C |
|                             |           | ctc                                |       |
| SpAiQ <sub>n</sub> BseR1 5' | (5' – 3') | tcagagcctgaaagagaggagtgatccgcagc   | 68 °C |
|                             |           | ag                                 |       |
| SpAiQ <sub>n</sub> BseR1 3' | (5' – 3') | ctgctgcggatcactcctctctttcaggctctga | 68 °C |

##### **SpA-idpolyQ construction and amplification**

|                              |           |                              |       |
|------------------------------|-----------|------------------------------|-------|
| SpAidQ <sub>n</sub> Δstop 5' | (5' – 3') | gcggcccggggcaggatccgaattc    | 67 °C |
| SpAidQ <sub>n</sub> Δstop 3' | (5' – 3') | gaattcggatcctgccccgggcccgc   | 67 °C |
| SpAampBamH1 5'               | (5' – 3') | cagggccacgtgggatccgataacaaat | 63 °C |

|                             |           |                                     |       |
|-----------------------------|-----------|-------------------------------------|-------|
| SpAampXho1 3'               | (5' – 3') | ctcgaattcctcgagtcacccgggcccgcggtt   | 71 °C |
| T7 prom 5'                  | (5' – 3') | As above                            |       |
| SpAidQ <sub>n</sub> Bsg1 3' | (5' – 3') | ttaaatttgttatcggtgcaggatcctgccccggg | 67 °C |

### SpA<sub>2</sub>-polyQ construction

|  |           |   |       |
|--|-----------|---|-------|
| SpA <sub>2</sub> -Q <sub>n</sub> Kpn1ins5' | (5' – 3') | gccacgtggacgtcggtaccgataacaaatttaa<br>taa | 64 °C |
| SpA <sub>2</sub> -Q <sub>n</sub> Kpn1ins3' | (5' – 3') | ttattaaatttgttatcggtaccgacgtccacgtgg<br>c | 64 °C |
| SpA <sub>2</sub> Q <sub>n</sub> Pml1amp5'  | (5' – 3') | gggcgccagggccacgtggacgtcgataac            | 71 °C |
| SpA <sub>2</sub> Q <sub>n</sub> Kpn1amp3'  | (5' – 3') | cccgggccgcgggtaccttcggcgctg               | 75 °C |

### SpA<sub>3</sub>-polyQ construction

|  |           |                                |       |
|--|-----------|--------------------------------|-------|
| SpA <sub>3</sub> Q <sub>n</sub> Spe1ins 5' | (5' – 3') | cacgtggacgtcactagtataacaaattta | 59 °C |
| SpA <sub>3</sub> Q <sub>n</sub> Spe1ins 3' | (5' – 3') | taaatttgttatcactagtacgtccacgtg | 59 °C |
| SpA <sub>2</sub> Q <sub>n</sub> Pml1amp 5' | (5' – 3') | As above                       |       |
| SpA <sub>2</sub> Q <sub>n</sub> Spe1amp 3' | (5' – 3') | cccgggccgcggactagtttcggcgctg   | 72 °C |

## 2.2.4 Polymerase Chain Reaction (PCR)

Amplification of DNA was performed by the polymerase chain reaction (PCR) in a Corbin Research Thermal Cycler using the oligonucleotides in Section 2.2.3. *Pfu* DNA polymerase (Promega) was used for all PCR reactions.

### General PCR conditions

|                                  | Stock     | Final Concentration | Volume   |
|----------------------------------|-----------|---------------------|----------|
| <b>Pfu DNA polymerase buffer</b> | 10x       | 1x                  | 2.5 µL   |
| <b>dNTP's</b>                    | 2 mM      | 200 µM              | 2.5 µL   |
| <b>Primers (5' and 3')</b>       | 10 µM     | 500 nM              | 1.25 µL  |
| <b>DNA template</b>              | 400 ng/µL | 10-500 ng           | 0.5 µL   |
| <b><i>Pfu</i></b>                | 3 u/µL    | 3 u                 | 0.5 µL   |
| <b>ddH<sub>2</sub>O</b>          | -         | -                   | to 25 µL |

### Temperature cycling

| Step                        | Temperature | Time     | Cycles |
|-----------------------------|-------------|----------|--------|
| <b>Initial Denaturation</b> | 95°C        | 1 min    | 1      |
| <b>Denaturation</b>         | 95°C        | 1 min    |        |
| <b>Annealing</b>            | X°C         | 1 min    | 30     |
| <b>Extension</b>            | 72°C        | 2 min/kb |        |
| <b>Final Extension</b>      | 72°C        | 2 min/kb | 1      |
| <b>Hold</b>                 | 25°C        | -        | -      |

X= 5 °C below the melting temperature of the primers

### **2.2.5 QuikChange™ site-directed mutagenesis**

The QuikChange™ (Stratagene) method was used to introduce mutations, restriction sites and short CAG segments within the template DNA. Firstly, complementary primers containing the mutation of interest were used to amplify the whole plasmid. As the parental DNA is methylated, it can be digested by *Dpn1*, leaving the unmethylated, mutated PCR product. The mutated DNA is subsequently transformed and the *E. coli* machinery repairs the nick.

#### *QuikChange™ PCR conditions*

|                                  | <b>Stock</b> | <b>Final Concentration</b> | <b>Volume</b> |
|----------------------------------|--------------|----------------------------|---------------|
| <b>Pfu DNA polymerase buffer</b> | 10x          | 1x                         | 2.5 µL        |
| <b>dNTPs</b>                     | 2 mM         | 200 µM                     | 2.5 µL        |
| <b>Primers (5' and 3')</b>       | 10 µM        | 320 nM                     | 1.25 µL       |
| <b>DNA template</b>              | 50 ng/µL     | 50 ng                      | 0.5 µL        |
| <b><i>Pfu</i></b>                | 3 u/µL       | 3 u                        | 0.5 µL        |
| <b>ddH<sub>2</sub>O</b>          | -            | -                          | to 25 µL      |

#### *Temperature Cycling*

| <b>Step</b>                 | <b>Temperature</b> | <b>Time</b> | <b>Cycles</b> |
|-----------------------------|--------------------|-------------|---------------|
| <b>Initial Denaturation</b> | 95°C               | 1 min       | 1             |
| <b>Denaturation</b>         | 95°C               | 1 min       |               |
| <b>Annealing</b>            | X°C                | 1min        | Y             |
| <b>Extension</b>            | 72°C               | 2 min/kb    |               |
| <b>Final Extension</b>      | 72°C               | 2 min/kb    | 1             |
| <b>Hold</b>                 | 25°C               | -           | -             |

X= 5 °C below the melting temperature of the primers

Y= The number of cycles were dependent on the nature of the insertion. 18 cycles were used for the introduction of single mutations and restriction sites and 20 cycles were used for insertion of short segments of CAG repeats.

#### *Dpn1 digest*

|                    |         |
|--------------------|---------|
| PCR product        | 20 µL   |
| Buffer (10x)       | 2.5 µL  |
| <i>Dpn1</i>        | 1 µL    |
| ddH <sub>2</sub> O | 1.25 µL |

Reactions were incubated at 37 °C for 2 h. 5 µL of the reaction mixture was then used to transform 50 µL of NovaBlue™ (Novagen) competent cells. Resulting colonies were screened (see section 2.2.13) to confirm the presence of the insertion.

### **2.2.6 Agarose Gel Electrophoresis**

A 50x TAE stock solution was prepared [242 g Tris, 57.1 mL glacial acetic acid, 100 ml 0.5 M EDTA, pH 8.0]. To 1x TAE, a final concentration of 1 % (w/v) or 2 % (w/v) of agarose (Promega) was added and dissolved by heating. The solution was cooled before pouring into a DNA gel frame with comb and allowed to set. The gel frame was then placed into the tank containing 1x TAE. The samples were mixed with 1 µL EZ-Vision™ One (Amresco) and loaded into the wells. Electrophoresis was performed at a constant voltage of 100 V until the dye front had reached the bottom. DNA was then visualized using UV light.

### **2.2.7 Restriction Enzyme Digestion**

Restriction enzyme digests of PCR products and plasmids were performed using enzymes and their respective buffers from Promega and New England Biolabs (NEB).

#### **General Reaction setup**

|                           |          |
|---------------------------|----------|
| DNA (PCR product/plasmid) | 30 µL    |
| Buffer (10x)              | 5 µL     |
| Enzyme                    | 0.5-5 µL |
| BSA (100x)                | 0.5 µL   |
| ddH <sub>2</sub> O        | To 50 µL |

Double digests were performed with compatible buffering systems. Digests were typically incubated for 2 hours at 37 °C, unless otherwise recommended for a specific enzyme.

### **2.2.8 Gel Purification**

To obtain purified DNA fragments prior to ligation, the fragment of interest was excised from an agarose gel and gel purification performed using the Wizard PCR and Gel purification kit (Promega) according to the manufacturer's instructions.

### **2.2.9 Ligations**

T4 DNA ligase and the 10x ligation buffer (Promega) were used for the ligation reactions. For cohesive end ligations, 1:3 or 1:5 molar ratios of vector to insert were used. The amount of insert required for each molar ratio was determined using the equation below:

$$\frac{\text{ng vector} \times \text{size insert (kb)}}{\text{size vector (kb)}} \times \text{molar ratio} = \text{ng insert} \quad \text{Eq. 2.1}$$

All ligation reactions were incubated at room temperature for 1.5 h.

### **2.2.10 Plasmid purification**

All expression plasmids contained genes conferring Amp resistance to host cells. 5 mL of 2YT media was inoculated with *E. coli* cells containing the plasmid of interest. The cells were grown overnight in a shaking incubator at 37 °C. Plasmid DNA was purified using a Wizard Plus SV Miniprep DNA purification system (Promega) according to the manufacturer's guidelines.

### **2.2.11 Preparation of competent cells**

Competent cells were prepared using the rubidium chloride method of Hanahan (197).

Transformation Buffer 1: 100 mM Rubidium Chloride, 50 mM Manganese Chloride, 10 mM Calcium Chloride, 30 mM Potassium acetate, 15 % (w/v) Glycerol, pH 5.8.

---

Transformation Buffer 2: 10 mM Rubidium Chloride, 75 mM Calcium Chloride, 10 mM MOPS, 15 % (w/v) Glycerol, pH 6.8.

Solutions were filter sterilized and stored at 4 °C.

A scraping of cells from a glycerol stock was added to 5 ml 2YT media and allowed to grow overnight at 37 °C. 500 µL of the starter culture was used to inoculate 50 mL of 2YT in a 250 mL flask and incubated at 37 °C under shaking conditions until the optical density at 600 nm (OD<sub>600</sub>) was approximately 0.5. The culture was then chilled on ice for 10 minutes. The cells were then pelleted at 3000 rpm for 10 min. The supernatant was discarded and the cells resuspended in 15 mL of Transformation buffer 1. The cells were allowed to incubate on ice for 15-30 min and then pelleted again. The cell pellet was then resuspended in 4 mL of Transformation buffer 2 and allowed to incubate on ice for 15 min. Cells were aliquoted into 100 µL volumes, snap frozen in a dry ice/ethanol bath and immediately transferred to -80 °C for storage.

### **2.2.12 Bacterial transformations**

1 µL of purified plasmid or 5 µL of PCR product/ligation product was added to 100 µL of competent *E. coli* cells and incubated on ice for 30 min. The cells were heat-shocked at 42 °C for 90 sec and then incubated on ice for 2 min to recover. When PCR products were transformed, cells were allowed to recover for 1 h in 150 µL of SOC media at 37 °C, 200 rpm. Cells were then plated onto 2YT agar plates containing 0.1 mg/mL Amp and incubated at 37°C for 12-16 h.

### **2.2.13 Colony screening**

Transformants were screened by restriction enzyme digestion in reactions where a novel restriction site was introduced. For confirmation of restriction site or point mutation

insertion, positive transformants of the reactions were submitted to Micromon DNA sequencing (Microbiology department, Monash University). Samples were set up as follows:

|                         |          |
|-------------------------|----------|
| DNA                     | 500 ng   |
| Primer                  | 0.5 pmol |
| Big Dye Terminator v3.1 | 1 µL     |
| 10x Taq Buffer          | 2 µL     |
| ddH <sub>2</sub> O      | To 20 µL |

#### *Thermal Cycling conditions*

| Step                        | Temperature | Time  | Cycle |
|-----------------------------|-------------|-------|-------|
| <b>Initial Denaturation</b> | 90 °C       | 1 min | 1     |
| <b>Denaturation</b>         | 90 °C       | 1 min |       |
| <b>Annealing</b>            | 50 °C       | 1 min | 25    |
| <b>Extension</b>            | 60 °C       | 4 min |       |
| <b>Final Extension</b>      | 60 °C       | 4 min | 1     |
| <b>Hold</b>                 | 25 °C       | -     |       |

Following the PCR, 1/10 volume of 3 M sodium acetate (pH 5.2) and 2 volumes of 100 % (v/v) ethanol (ice-cold) were added and samples were incubated at -80 °C for 20 min. The sample was centrifuged for 15 min at 13,000 rpm and the supernatant removed. ½ volume of 70 % (v/v) ethanol was added, with the sample then centrifuged for 5 min at 13,000 rpm. This wash step was repeated twice and any remaining ethanol in the sample was evaporated by incubation of samples on a dry heating block set at 90 °C for 5 min.

## **2.3 Preparation of Proteins**

### **2.3.1 Buffers/solutions**

Lysis buffer: 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 1 mM BME, 0.1 % (v/v) Triton X-100, pH 8.0.

Wash buffer: 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 1 mM BME, 0.1 % (v/v) Triton X-100, 20 mM imidazole, pH 8.0.

Elution buffer: 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 1 mM BME, 0.1 % (v/v) Triton X-100, 250 mM imidazole, pH 8.0.

Sodium acetate buffer: 20 mM NaCH<sub>3</sub>COOH, 100 mM NaCl, 10 % (v/v) glycerol, pH 5.0.

Tris(hydroxymethyl)aminomethane (Tris) buffered saline (TBS):  
100 mM TrisHCl, 80 mM NaCl, pH 7.4 (at 37 °C)

### **2.3.2 Expression of SpA-polyQ variants**

SpA-polyQ fusion proteins were expressed in BL21(DE3) cells by autoinduction (198). A 2YT agar plate with Amp (0.1 mg/mL) was inoculated with BL21(DE3) and incubated at 37 °C overnight. A scraping of cells was added to 200 mL of autoinduction media (Novagen), containing Amp (0.1 mg/mL), and cells were grown under shaking conditions at 37 °C for 6 h. Expression then occurred as cells were incubated for a further 16 h at 25 °C. Cells were harvested by centrifugation at 5,000 rpm for 10 min at 4 °C, resuspended in Lysis buffer and stored at -80 °C until purification.

### **2.3.3 Expression of ataxin-3 variants**

Ataxin-3(Q64) and Josephin variants in pExpLicHis were expressed in BL21(DE3) *E. coli* cells. 2YT supplemented with 0.1 mg/mL ampicillin was inoculated with these cells and grown for 12-16 h at 37 °C, 180 rpm. This culture was added to 2YT media at 1 % (v/v) and grown at 37 °C, 180 rpm until OD<sub>600</sub> ~ 0.6. Protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM and the bacterial cultures



incubated for a further 4 h at 28 °C, 180 rpm. Cells were harvested by centrifugation at 5,000 rpm and pellets resuspended in lysis buffer and stored at -80 °C.

### **2.3.4 Expression of $^{15}\text{N}$ and $^{15}\text{N}/^{13}\text{C}$ -labeled proteins for NMR spectroscopy**

Labeled proteins for NMR experiments uniformly enriched with either  $^{15}\text{N}$  only or  $^{15}\text{N}/^{13}\text{C}$  were expressed according to the protocol of Marley *et al.* (199).

#### *Minimal media*

To prepare 1L of minimal media

|  | Volume (mL) |
|--|-------------|
| 5 x M9 salts   | 200         |
| D-glucose stock (20 g/100 mL) (0.2 µm filter sterilized) | 20          |
| Basal Vitamins Eagle Media                               | 10          |
| 1 M $\text{MgSO}_4$ (autoclaved)                         | 2           |
| 1 M $\text{CaCl}_2$ (autoclaved)                         | 0.1         |

To prepare 5 X M9 Salts (autoclaved) (1 L)

|   |       |
|---|-------|
| $\text{KH}_2\text{PO}_4$  | 15 g  |
| $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$             | 64 g  |
| $\text{NaCl}$   | 2.5 g |
| $\text{NH}_4\text{Cl}$ (pH adjusted to 7.2 with $\text{NaOH}$ ) | 5 g   |

For isotopic labeling,  $^{15}\text{NH}_4\text{Cl}$  was substituted for unlabeled  $\text{NH}_4\text{Cl}$  and  $^{13}\text{C}$  D-glucose was substituted for unlabeled D-glucose.

An overnight culture was prepared by inoculation of 50 mL 2YT containing Amp (0.1 mg/mL). 1 L of 2YT (containing Amp (0.1 mg/ml) was inoculated with 10 mL of the overnight culture and cells were grown to an  $\text{OD}_{600}$  of 0.7 in 2YT media at 37 °C, 180 rpm. Cells were pelleted by centrifugation for 10 min at 5,000 rpm. The cell pellet was washed by resuspension in 100 mL of M9 minimal salts followed by centrifugation at 4,000 rpm. The supernatant was discarded and the wash step repeated. The cell pellet

was then resuspended in 250 mL of minimal media (containing 0.1 mg/mL Amp) and added to a 2 L sterile, baffled flask. Cells were incubated at 28 °C, 180 rpm for 1 h to allow the cells to recover prior to induction with IPTG to a final concentration of 0.5 mM. Cells were incubated at 28 °C, 180 rpm for a further 4 h, harvested and pellets resuspended in lysis buffer and stored at -80 °C.

### **2.3.5 Purification of SpA-polyQ proteins**

Cell lysates were thawed and incubated with lysozyme (1.4 mM) and PMSF (50 µM) for 20 min. The cell lysate was then sonicated and centrifuged for 30 min at 20,000 rpm at 4 °C. The pellet was discarded and the supernatant collected and filtered through a 0.45 µm filter. The lysate was then loaded onto a pre-equilibrated 5 mL HisTrap HP column (Amersham) at a flow rate of 5 mL/min. After loading, the column was washed with 10 column volumes of wash buffer using an Akta FPLC system (Amersham). The protein was eluted from the column using an imidazole gradient from 20 mM (wash buffer) to 250 mM (elution buffer) over ten column volumes. 1 mL fractions were collected, run on SDS-PAGE and appropriate fractions were pooled and loaded onto either a Hiload Superdex™ 200 16/60 prep grade column (for SpAcQ40 and larger variants) or a Hiload Superdex™ 75 16/60 prep grade (Pharmacia) (for smaller variants) that had been pre-equilibrated with sodium acetate buffer prior to loading. Fractions were collected, analyzed by SDS-PAGE, filtered through a 0.45 µm filter and stored at -80 °C. High yields of protein were obtained at >95 % purity.

### **2.3.6 Purification of ataxin-3 variants**

The cell lysate was thawed and incubated with lysozyme (1.4 mM), DNase 1 (0.1 mg/mL) and Pefabloc (0.5 mg/mL) for 20 min at 4 °C. Cells were then sonicated and

further Pefabloc (0.5 mg/mL) added. Lysed cells were centrifuged at 20,000 rpm for 30 min at 4 °C. The pellet was discarded and the supernatant collected and filtered through a 0.45 µm filter. 10 % (v/v) p-aminobenzamidine agarose was then added to the lysate and incubated on a rotary orbitor for 1 h at 4 °C. The flow through was collected and Pefabloc (0.5 mg/mL) and imidazole (10 mM) were added. This solution was loaded onto a 5 mL HisTrap HP column (Amersham) pre-equilibrated with wash buffer. The column was washed with 10 column volumes of wash buffer using an Akta FPLC system (Amersham), and then eluted with an imidazole gradient from 20 mM (wash buffer) to 250 mM (elution buffer). 1 mL fractions were collected, analyzed by SDS-PAGE and appropriate fractions were pooled and Pefabloc (0.5 mg/mL) added. The protein was then loaded onto a Hiload Superdex™ 200 prep grade column that was pre-equilibrated with TBS. Fractions were collected as they eluted and the presence of purified protein verified by SDS-PAGE. Protein was obtained at >95 % purity. The protein was concentrated using an Amicon Ultra (Millipore) centrifuge concentrator with a 10 kDa molecular weight cutoff. The protein was filtered through a 0.45 µm filter, aliquoted and stored at -80 °C.

### **2.3.7 Preparation of αB-crystallin**

Lyophilized αB-crystallin was a gift of the laboratory of Professor John Carver.

### **2.3.8 Determination of Protein concentration**

The concentration of SpA-polyQ proteins was determined by measurement of the absorbance at 280 nm (Abs<sub>280</sub>) and verified by measuring the intrinsic fluorescence of proteins unfolded in GuSCN. The concentration of ataxin-3 variants was determined by measurement of Abs<sub>280</sub> and a bicinchoninic acid (BCA) assay (Pierce). The BCA assays were performed according to the manufacturer's instructions using bovine serum

albumin (BSA) as a protein standard. Extinction coefficients ( $\epsilon$ ) for all SpA-polyQ and ataxin-3 variants were calculated by the Equation 2.2 (200, 201):

$$\epsilon = \text{Number (Tyrosine)} \times \text{Ext (Tyrosine)} + \text{Number (Tryptophan)} \times \text{Ext (Tryptophan)} + \text{Number (Cysteine)} \times \text{Ext (Cysteine)} \quad \text{Eq 2.2}$$

Where Ext (Tyrosine) = 1490, Ext (Tryptophan) = 5500, Ext (Cysteine) = 125.

## **2.4 SDS-PAGE**

12 % Tris-tricine SDS-PAGE or 15 % Tris-glycine SDS-PAGE were used for the analysis of SpA-polyQ variants. 12.5 % Tris-glycine SDS-PAGE was used for the analysis of ataxin-3 variants.

### *Solutions required for preparation of Tris-glycine gels*

|                         |  |
|-------------------------|--|
| 4x running gel buffer:  | 1.5 M TrisHCl, pH 8.8.                                     |
| 4x stacking gel buffer: | 0.5 M TrisHCl, pH 6.8.                                     |
| SDS Tank buffer:        | 0.025 M TrisHCl, 0.192 M glycine, 0.1 % (w/v) SDS, pH 8.3. |

### *Solution required for preparation of Tris-tricine gels*

|                 |   |
|-----------------|---|
| Solution A:     | 3 M TrisHCl, 0.3 % (w/v) SDS, pH 8.45.                  |
| Solution B:     | 49.5 % (w/v) Acrylamide, 3 % (w/v) Bis-acrylamide.      |
| Anode Buffer:   | 0.2 M TrisHCl, pH 8.9.                                  |
| Cathode Buffer: | 0.1 M TrisHCl, 0.1 M Tricine, 0.1 % (w/v) SDS, pH 8.25. |

### *General SDS-PAGE solutions*

|                    |  |
|--------------------|--|
| 2x Loading buffer: | 0.125 M TrisHCl, 20 % (v/v) glycerol, 4 % (w/v) SDS, 0.2 M DTT, 0.02 % (w/v) Bromophenol Blue, pH 6.8. |
|--------------------|--|

Stain solution: 40 % (v/v) methanol, 7 % (v/v) acetic acid, 0.025 % (w/v) coomassie brilliant blue (R250).

Destain solution: 40 % (v/v) methanol, 7 % (v/v) acetic acid.

**Table 2.1. Components of two Tris-glycine SDS-PAGE gels**

|                              | <i>Running Gel</i> |            | <i>Stacking Gel</i> |
|------------------------------|--------------------|------------|---------------------|
|                              | 12.5 % (v/v)       | 15 % (v/v) |                     |
| Bis-acrylamide(monomer) (mL) | 4.12               | 4.94       | 0.44                |
| 4x Running buffer (mL)       | 2.50               | 2.50       | -                   |
| 4x Stacking buffer (mL)      | -                  | -          | 0.83                |
| 10 % (w/v) SDS (μL)          | 100                | 100        | 33                  |
| ddH <sub>2</sub> O (mL)      | 3.18               | 2.36       | 2.03                |
| 25 % (w/v) APS (μL)          | 60                 | 60         | 20                  |
| TEMED (μL)                   | 4                  | 4          | 3                   |

**Table 2.2. Components of two 12 % Tris-tricine SDS-PAGE gels**

|                         | <i>Running Gel</i> | <i>Stacking Gel</i> |
|-------------------------|--------------------|---------------------|
| Solution A              | 3                  | 0.75                |
| Solution B              | 2.16               | 0.25                |
| ddH <sub>2</sub> O (mL) | 4.7                | 2                   |
| 25 % (w/v) APS (μL)     | 60                 | 20                  |
| TEMED (μL)              | 4                  | 3                   |

SDS gels were prepared according to Tables 2.1 and 2.2 and run using a Biorad minigel system. For tris-glycine gels SDS tank buffer was present in both the anode and cathode chambers, whilst tris-glycine gels utilized different anode and cathode buffers (described above). Samples were diluted in 2x loading buffer, boiled for 5 min and applied to the wells of the stacking gel. Gels were run for 40 min at a constant current of 40 mA. Gels were then stained with stain solution.

## **2.5 Size exclusion chromatography**

Size exclusion chromatography (SEC) was carried out on an Akta FPLC (Amersham). The elution profiles of SpA-polyQ proteins were analyzed using a Superdex™ 75 10/300 gel filtration column (Amersham), pre-equilibrated with sodium acetate buffer

and elution monitored by Abs<sub>280</sub> detection. In experiments monitoring the effects of  $\alpha$ B-crystallin on the aggregation of ataxin-3 variants, samples were taken at indicated times and insoluble aggregates removed by centrifugation at 13,000 rpm. The supernatant was injected onto a Superose™ 12 HR 10/30 column (Amersham) that had been pre-equilibrated with TBS.

## **2.6 Circular dichroism**

Far-UV circular dichroism (CD) studies were performed on a Jasco 810 spectropolarimeter thermostatted by a Jasco peltier. For secondary structure analysis, measurements were performed using a 0.01 cm pathlength quartz cuvette. Spectra were measured from 190-260 nm using a scan speed of 50 nm/min and three scans accumulated for each sample. Far-UV CD measurements were converted to molar ellipticity ( $\theta$ ) using equation 2.3:

$$[\theta]_{MW} = \frac{\theta_{obs}}{10 \times l \times c} \quad \text{Eq. 2.3}$$

Where  $l$  is the pathlength (cm),  $C$  is the protein concentration (M) and  $\theta_{obs}$  is the CD signal in millidegrees. The units for molar ellipticity are mdeg.cm<sup>2</sup>.dmol<sup>-1</sup>. To estimate the secondary structural contributions, the SpA-polyQ far-UV CD spectra were deconvoluted using the K2D (202) algorithm available at the online DICHROWEB facility (203).

## **2.7 Thermal and guanidine denaturation**

### **2.7.1 Thermal denaturation**

SpA-polyQ proteins at a concentration of 20  $\mu$ M in sodium acetate buffer in the presence of 1 M GuHCl were denatured as the temperature was increased from 40  $^{\circ}$ C to 90  $^{\circ}$ C at a rate of 1  $^{\circ}$ C/min. Denaturation was performed in a 0.1 cm path length stoppered cuvette to eliminate evaporation. The change in CD signal at 222 nm was monitored.

### **2.7.2 GuHCl unfolding**

Stock solutions of GuHCl were prepared in sodium acetate buffer. Equilibrium unfolding titrations were set up with GuHCl ranging from 0 – 6 M, with a constant protein concentration of 20  $\mu$ M. Proteins were incubated in GuHCl solutions for 30 min at room temperature and the CD signal at 222 nm ( $CD_{222}$ ) was monitored. Reversibility of denaturation was verified by performing refolding studies, in which the protein was initially denatured in 6 M GuHCl followed by titration into the 0 – 6 M GuHCl solutions. Single transitions for un/folding were observed and data were fit to a two-state unfolding model described by Santoro and Bolen (204):

$$I = \frac{(I_f + m_f[U]) + (I_u + m_u[U]) \exp(-(\Delta G^{\circ} - m[U])/(RT))}{1 + \exp(-(\Delta G^{\circ} - m[U])/(RT))} \quad \text{Eq 2.4}$$

Where  $I$  is the measured fluorescence signal,  $T$  is the temperature in degrees Kelvin,  $R$  is the gas constant (1.98 cal/mol/K),  $I_f$  is the value of the folded signal in the absence of GuHCl,  $m_f$  is the slope of the GuHCl dependence,  $I_u$  is the value of the unfolded signal in the absence of GuHCl,  $m_u$  is the

slope of the dependence of the fluorescence signal on GuHCl for the unfolded protein,  $[U]$  is the GuHCl concentration,  $\Delta G^0$  is the free energy of unfolding at 0 M GuHCl, and  $m$  is the dependence of the free energy of unfolding on the denaturant concentration. The midpoint of the transition could be determined by fitting the data to this model.

## **2.8 Protein Aggregation assays**

### **2.8.1 SpA-polyQ variants**

The SpA-polyQ proteins were incubated at a concentration of 20  $\mu$ M, 50  $\mu$ M or 65  $\mu$ M in 20 mM sodium acetate buffer (pH 5.0) with 2 mM PMSF. For discontinuous assays, samples were incubated at 37 °C without shaking in airtight containers to avoid evaporation. At various timepoints, samples were removed and assayed as described in the relevant methods sections.

### **2.8.2 Ataxin-3 variants**

Ataxin-3 variants were incubated at a concentration of 20  $\mu$ M in TBS (pH 7.4) buffer containing 2 mM PMSF, 5 mM EDTA and 15 mM BME. Samples were incubated at 37 °C under non-shaking conditions in airtight containers to eliminate evaporation. At various timepoints, samples were removed and assayed as described in the relevant methods sections.

### **2.8.3 Aggregation assays with the addition of $\alpha$ B-c**

Aggregation assays containing SpA-polyQ or Ataxin-3 variants were set up as described above. Variable concentrations of  $\alpha$ B-c were added to the samples, ranging from 2  $\mu$ M to 40  $\mu$ M.



---

**2.8.4 QBP1 assays**

The QBP1 peptide, with the amino acid sequence, SNWKWWPGIFD (118) was purchased from AusPep. The peptide was obtained in a lyophilized form that was solubilized in a solution of 50 % dimethyl sulfoxide (DMSO)/50 % sodium acetate buffer to a concentration of 8 mM. QBP1 was incubated with SpA-polyQ aggregation assays, as described above, at a final concentration of 120  $\mu$ M. Control assays with identical final DMSO concentrations were performed and showed no significant effect of DMSO addition.

**2.8.5 ThioflavinT fluorescence****2.8.5.1 Discontinuous ThT measurements**

Samples were taken from the aggregation assays (described above) and diluted in buffer containing 20  $\mu$ M ThT. Samples were placed in quartz fluorescence cuvettes (Starna), and the emission intensity at 480 nm monitored ( $\lambda_{\text{ex}} = 445$  nm).

**2.8.5.2 Continuous ThT measurements**

ThT fluorescence was monitored using a BMG laboratories Fluostar Optima fluorescence platereader. Continuous assays were set up as described for the aggregation assays above, however, 20  $\mu$ M ThT was present for the reaction timecourse. Excitation and emission filters of 450 and 490 nm were used, respectively. The reactions were performed in black 384-well clear bottom plates (Corning) and excitation and emission read from the bottom. All reactions were incubated at 37 °C, without shaking, and the top of the plate was sealed to eliminate evaporation. The aggregation kinetics obtained from continuous plate reader assays (performed in the

presence of ThT) were comparable to kinetic data obtained from discontinuous ThT measurements.

## **2.9 Membrane filter trap assay**

### *Solutions*

|                        |  |
|------------------------|--|
| Stability Buffer:      | 4 % (w/v) SDS, 100 mM DTT.   |
| Blocking Buffer:       | 5 % skim milk powder in blotting buffer  |
| Blotting Buffer:       | 20 mM TrisHCl, 0.2 M NaCl, pH 7.4.   |
| Tween Blotting Buffer: | 20 mM TrisHCl, 0.2 M NaCl, 0.5 % (v/v) Tween, pH 7.4.  |
| Primary Antibody:      | Mouse anti-histidine tag (Serotec), 1:5000 dilution in blotting buffer containing 0.5 % (w/v) BSA.                             |
| Secondary Antibody:    | Horseradish peroxidase (HRP)-conjugated anti-mouse (Amersham), 1:30000 dilution in blotting buffer containing 0.5 % (w/v) BSA. |

Protein aliquots were diluted into stability buffer at a 1:1 (v/v) ratio. The samples were heated for 5 min at 100 °C and then diluted to 200 µL with 2 % (w/v) SDS. 100 µL of each sample was filtered through a pre-equilibrated cellulose acetate membrane (Schleicher and Schuell, 0.2 µm) using a Bio-Dot SF microfiltration unit (Bio-Rad). The membrane was washed twice by filtering through 150 µL of 0.1 % (w/v) SDS. The membrane was then soaked in blocking buffer for 1.5 h and then rinsed with Tween blotting buffer. The membrane was soaked for 1.5 h in a solution of the antihistidine primary antibody with blocking buffer containing 0.5 % (w/v) BSA. The membrane was washed in Tween blotting buffer (3 x 10 min) and incubated for 1.5 h in the secondary antibody with blotting buffer containing 0.5 % (w/v) BSA. The membrane was then washed again with Tween blotting buffer (3 x 10 min). The blot was treated with ECL reagents for 1 min and then exposed onto film for 10 s, 30 s and 2 min in a

Hypercassette™ (Amersham), which was then developed in a Fuji Processor FPM-100A.

## **2.10 Transmission electron microscopy**

Samples were adsorbed onto carbon grids and negatively stained with 1 % (w/v) aqueous uranyl acetate (pH 8.0). Samples were then analyzed by transmission electron microscopy on a Hitachi H7500 electron microscope using an acceleration voltage of 80 kV.

## **2.11 Nuclear Magnetic Resonance (NMR) Spectroscopy**

Unlabelled,  $^{15}\text{N}$ -labeled and  $^{15}\text{N}/^{13}\text{C}$ -labeled proteins for NMR analysis were prepared as described in section 2.3. All experiments were recorded on a Varian Unity 600 MHz spectrometer equipped with a triple resonance cold probe.

Triple resonance experiments were performed to obtain assignments for SpA-cQ5 and Josephin (205). 3D HNCO, HNCA and CBCA(CO)NH spectra were acquired using  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled samples. To obtain interproton distances two- and three-dimensional nuclear overhauser enhancement spectroscopy (NOESY) experiments were recorded (206).

$^{15}\text{N}$ - $^1\text{H}$ -heteronuclear single-quantum coherence (HSQC) experiments spectra were collected for SpA-cQ5, SpA-cQ52 and Josephin. To determine the  $\alpha\text{B-c}$  interaction surface of Josephin, samples of Josephin were titrated with increasing concentrations of  $\alpha\text{B-c}$  and  $^{15}\text{N}$ - $^1\text{H}$ -HSQC spectra were acquired.  $\alpha\text{B-c}$ :Josephin ratios of 0, 0.1, 0.25, 0.6, 2 and 5 were used.

Paramagnetic relaxation enhancement (PRE) was used to determine the surface exposed residues of Josephin involved in the interaction with  $\alpha\text{B-c}$ . The gadolinium-

based paramagnetic relaxation agent, Gd-diethylenetriamine pentaacetic acid-bismethylamide (Gd(DTPA-BMA)) at a concentration of 5 mM, was added to a 0.3 mM  $^{15}\text{N}$ -labelled Josephin sample.  $^{15}\text{N}$ -HSQC spectra were recorded in the absence and presence of 0.6 mM  $\alpha\text{B-c}$ .

All data were processed using nmrPipe with sine squared window functions, zero filling and linear prediction in the indirect dimensions. Spectra were visualized and analyzed using Sparky (Goddard and Keller, UCSF). NMR experiments were performed in collaboration with Martin Scanlon, James Horne and Stephen Headey (Monash Institute of Pharmaceutical Sciences, Monash University).

## **Chapter 3**

### **A cyclic method for the incremental expansion of polyglutamine repeats in recombinant proteins**

Submitted as a ‘notes and tips’ article to *Analytical Biochemistry*.

---

## **A cyclic method for the incremental expansion of polyglutamine repeats in recombinant proteins**

Amy L. Robertson, Lisa D. Cabrita and Stephen P. Bottomley.

Department of Biochemistry and Molecular Biology, Monash University, Victoria,  
Australia.

Correspondence should be addressed to: Stephen P. Bottomley, Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia.  
Tel: 61-3-99029362; Email: [steve.bottomley@med.monash.edu.au](mailto:steve.bottomley@med.monash.edu.au)

### **Abstract**

The polyglutamine diseases are caused by the expansion of CAG repeats. A key step in understanding the disease mechanisms, at the DNA and protein level is the ability to produce recombinant proteins with specific length glutamine tracts. This is a time-consuming first step in setting up model systems to study the effects of polyglutamine expansion. Described here is a cyclic, PCR-based method for the amplification of CAG repeats, which we used to incrementally extend CAG length by 3-5 repeats per cycle. This method could be translated into various contexts where amplification of repeating elements is necessary.

Nine hereditary diseases have been found to correlate with the expansion of a CAG trinucleotide repeat, encoding glutamine, at a specific site within the open reading frame of a disease gene (reviewed in refs [1; 2]). These disorders are collectively termed the polyglutamine diseases, and include Huntington's disease and a number of spinocerebellar ataxias [1; 2]. Individuals with a CAG repeat number above 35-40 display a neurodegenerative phenotype, with deposition of protein aggregates as neuronal inclusions [3]. Investigating various aspects of CAG and glutamine repeats at the DNA and protein levels respectively, are important in clarifying the molecular mechanisms leading to pathogenesis in these diseases.

Intrinsic difficulties in the cloning and expression of proteins with variable length polyglutamine repeats have rate-limited research progress. Further, the availability of, and potential to produce constructs of specific repeat lengths has limited the breadth of studies. Most groups have studied the disease-causing proteins, amplifying constructs from disease genes. More recently a number of studies have utilized polyglutamine-containing model proteins to analyse disease-associated pathways [4; 5; 6; 7]. In many of these model systems, the polyglutamine segment was also PCR amplified from a disease gene, in addition to adjacent nucleotide regions [8; 9], necessary because of the non-specificity of primer binding to the repeating CAG motif. Recent evidence indicates that residues surrounding a repeat tract can modulate polyglutamine-dependent behaviour [10; 11; 12]. As such, the presence of extraneous amino acids has affected the interpretability of model protein studies.

Cassette insertion techniques have also been used for the insertion of synthetically prepared polyglutamine segments [13; 14; 15]. Using these methods there are issues with

annealing specificity, which decreases as the number of repeats in a homogeneous CAG tract becomes larger. Homogeneous CAG repeats are very unstable, and commonly contract in repeat number [16; 17]. This phenomenon is particularly prevalent in bacterial cells, and a number of groups have taken advantage of this for the uncontrolled amplification of homogeneous repeats [16]. Control of repeat number is forfeited, and large expansions in repeat lengths have been observed [17]. *In vivo* amplification methods are therefore not ideal if small, incremental or specific changes in repeat number are required.

Here, we describe the construction of a polyglutamine-containing model system, for which we have used a novel, cyclic, PCR-based amplification method to modulate the repeat number of a homogeneous CAG repeat tract. Using this method, we have relative control of the repeat length, and have incrementally expanded the repeat number by 3-5 glutamines per cycle.

A dual-fragment amplification and ligation method for the extension of polyglutamine repeats in Huntingtin exon 1 has been previously described by Peters and Ross [18]. This strategy involves the concurrent amplification of two overlapping fragments, both containing the complete polyglutamine-encoding region. The two fragments are digested with a Type 11S restriction endonuclease, enzymes for which the cleavage position is located outside of the recognition sequence (reviewed in [19]), reducing PCR-based problems associated with introducing restriction sites juxtaposed to a repeating DNA segment. The coordinated positioning of the unique Type 11S sites leaves fragments with ‘sticky ends’ located at either the 5’ or 3’ end of the CAG repeat. Subsequent fragment and vector ligation effectively doubles the trinucleotide repeat



number. This method, however, has limited application to all polyglutamine-containing systems, with efficiency dependent upon repeat location, vector design, and fragment size.

We extended on the approach described in [18] to construct a polyglutamine-containing model protein system, comprising the *Staphylococcus aureus* Protein A B domain (SpA), located N-terminal to various length polyglutamine (Qn) repeats (Figure 1). The series of proteins are herein referred to as SpA-cQn. This approach also utilizes the coordinated positioning of Type 11S recognition sites. The significant difference is that this method involves an initial once-off site-directed mutagenesis step for the introduction of a *BseR1* recognition sequence 7 bp upstream of the first CAG codon (Reaction 1, Figure 1). Subsequent digestion with *BseR1* and *Nde1*, for which the unique restriction site is located upstream of the SpA coding sequence, cleaves out the SpA gene and leaves a Q(n-1) repeat, and a TC overhang on the antisense strand (Figure 1). The end product of Reaction 1 is thus a linearized vector comprising Q(n-1) repeats.

In Reaction 2 (Figure 1) the entire SpA-cQn coding region is PCR-amplified using a 5' sequence specific primer, and a template-specific primer introducing a *Bsg1* recognition sequence 14bp downstream from the final CAG repeat. Cleavage with *Bsg1* leaves Q(n-1) repeats, and an AG overhang at the 3' end of the sense strand. The digested fragment can then be subcloned into the Q(n-1) linearized vector from Reaction 1, creating a construct comprising Q(2n-1) repeats. This product is then used as input for Reaction 2, with subsequent rounds proceeding in a cyclic process, whereby the length of the polyglutamine repeat is extended by Q(n-1) repeats in each cycle.

This approach can be used to rapidly and efficiently amplify the number of polyglutamine repeats in a controlled manner. We used this method to amplify the number

of CAG repeats in a model system comprising variable length C-terminal polyglutamine repeats. Failure to amplify the CAG repeat length using the approach described in [18] was attributed to the terminal location of the repeat. The single fragment amplification and ligation method described here overcomes the problems associated with this.

The original template for Reactions 1 and 2 was a synthetically prepared construct with 5 CAG repeats (SpA-cQ5). Upon introduction of the *BseR1* site and restriction enzyme digestion the product of Reaction 1 was a linearized vector containing 4 CAG repeats and an antisense TC overhang. This vector was used as a reactant for a number of ligation reactions with products from Reaction 2, yielding the SpA-cQ9 and SpA-cQ13 constructs. Concurrently, to speed up the amplification process, a *BseR1* site was introduced into the SpA-cQ13 construct, undergoing Reaction 1, yielding a linearized Q12 vector. This was then combined with the digested SpA-cQ9 and SpA-cQ13 fragments to produce SpA-cQ20 and SpA-cQ24. A cyclic process was then used to make constructs up to a repeat length of 62 glutamines (Figure 2a). Proteins comprising repeat lengths between 5 and 52 glutamines were successfully expressed and purified [20]. The SpA-cQ62 protein was not expressed in *E. coli* cells, probably due to the rapid aggregation propensity and toxicity of this protein. SpA-cQn proteins with 35 glutamines or greater formed SDS-insoluble aggregates [20].

We found that upon ligation, some repeat lengths deviated from the anticipated size (Figure 2). This is probably a result of the intrinsically unstable behaviour of CAG repeats [21]. The genetic instability implicated in the polyglutamine diseases is reflected *in vitro*, as the repeating nature of the DNA lends to the formation of irregular conformations, with basepair misalignment causing strand slippage and the formation of hairpin structures,

resulting in fluctuations in repeat length [21; 22]. The small repeat length fluctuations in this case were manageable, with DNA sequencing validating the CAG repeat number.

Compared to the method described in [18], in which periodic CAA repeats were inserted into the repeating CAG element, the tract in this case is maintained as a homogeneous CAG trinucleotide repeat. In this model system trinucleotide homogeneity allows for the possibility of frameshift modulation of the repeat tract to form polyalanine or polyserine segments. The method described here, however, is similarly adaptable to the incorporation of alternative codons, as in [18], and we have utilized this method to insert cysteine codons at specific locations within the polyglutamine tract.

The cyclic repeat amplification method described here was efficient in constructing a polyglutamine model protein system. This method has benefits over other approaches for cloning polyglutamine repeats. Compared with insertion techniques, this approach does not introduce extraneous amino acids in the regions flanking the repeat. The method described here builds upon that previously described in [18], reducing the dual fragment amplification and ligation strategy to a single fragment method. This has proven to have greater efficiency in constructing this model system. This method offers an efficient alternative for the controlled amplification of polyglutamine repeats and has potential for use in variable cloning applications, particularly in cases where repeating DNA motifs hinder the introduction of restriction sites directly adjacent to, or within a repeat.

**References**

- [1] R.L. Margolis, and C.A. Ross, Expansion explosion: new clues to the pathogenesis of repeat expansion neurodegenerative diseases. *Trends in Molecular Medicine* 7 (2001) 479-482.
- [2] H. Zoghbi, and H. Orr, Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* 23 (2000) 217-247.
- [3] J.F. Gusella, and M.E. MacDonald, Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nature Review Neuroscience* 1 (2000) 109-115.
- [4] M. Tanaka, I. Morishima, T. Akagi, T. Hashikawa, and N. Nukina, Intra- and Intermolecular beta-pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases. *J. Biol. Chem.* 276 (2001) 45470-45475.
- [5] A.G. Ladurner, and A.R. Fersht, Glutamine, Alanine or Glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* 273 (1997) 330-337.
- [6] Z. Ignatova, and L.M. Gierasch, Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein. *J Biol Chem* 281 (2006) 12959-67.
- [7] Y. Nagai, T. Inui, H.A. Popiel, N. Fujikake, K. Hasegawa, Y. Urade, Y. Goto, H. Naiki, and T. Toda, A toxic monomeric conformer of the polyglutamine protein. *Nat Struct Mol Biol* (2007).
- [8] L. Masino, G. Kelly, K. Leonard, Y. Trottier, and A. Pastore, Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Letters* 513 (2002) 267-272.
- [9] Y. Nagai, T. Tucker, H. Ren, D.J. Kenan, B.S. Henderson, J.D. Keene, W.J. Strittmatter, and J.R. Burke, Inhibition of polyglutamine protein aggregation and cell death by novel peptides identified by phage display screening. *J Biol Chem* 275 (2000) 10437-42.
- [10] K. Nozaki, O. Onodera, H. Takano, and S. Tsuji, Amino acid sequences flanking polyglutamine stretches influence their potential for aggregate formation. *Neuroreport* 12 (2001) 3357-3364.
- [11] A. Bhattacharyya, A.K. Thakur, V.M. Chellgren, G. Thiagarajan, A.D. Williams, B.W. Chellgren, T.P. Creamer, and R. Wetzel, Oligoproline effects on polyglutamine conformation and aggregation. *J Mol Biol* 355 (2006) 524-35.
- [12] M.G. Burke, R. Woscholski, and S.N. Yaliraki, Differential hydrophobicity drives self-assembly in Huntington's disease. *Proc Natl Acad Sci U S A* 100 (2003) 13928-33.
- [13] Y.W. Chen, Site-specific mutagenesis in a homogeneous polyglutamine tract: application to spinocerebellar ataxin-3. *Protein Eng* 16 (2003) 1-4.
- [14] M.K.M. Chow, A.M. Ellisdon, L.D. Cabrita, and S.P. Bottomley, Polyglutamine expansion in Ataxin-3 does not affect protein stability. *J. Biol. Chem.* 279 (2004) 47643-47651.
- [15] M.D. Tobelmann, R.L. Kerby, and R.M. Murphy, A strategy for generating polyglutamine 'length libraries' in model host proteins. *Protein Eng Des Sel* 21 (2008) 161-4.
- [16] K. Ohshima, S. Kang, and R.D. Wells, CTG triplet repeats from human hereditary diseases are dominant genetic expansion products in *Escherichia coli*. *J Biol Chem* 271 (1996) 1853-6.

- [17]P.S. Sarkar, H.C. Chang, F.B. Boudi, and S. Reddy, CTG repeats show bimodal amplification in *E. coli*. *Cell* 95 (1998) 531-40.
- [18]M.F. Peters, and C.A. Ross, Preparation of human cDNAs encoding expanded polyglutamine repeats. *Neurosci Lett* 275 (1999) 129-32.
- [19]R.J. Roberts, M. Belfort, T. Bestor, A.S. Bhagwat, T.A. Bickle, J. Bitinaite, R.M. Blumenthal, S. Degtyarev, D.T. Dryden, K. Dybvig, K. Firman, E.S. Gromova, R.I. Gumpert, S.E. Halford, S. Hattman, J. Heitman, D.P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T.R. Klaenhammer, I. Kobayashi, H. Kong, D.H. Kruger, S. Lacks, M.G. Marinus, M. Miyahara, R.D. Morgan, N.E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D.N. Rao, N. Reich, V.E. Repin, E.U. Selker, P.C. Shaw, D.C. Stein, B.L. Stoddard, W. Szybalski, T.A. Trautner, J.L. Van Etten, J.M. Vitor, G.G. Wilson, and S.Y. Xu, A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 31 (2003) 1805-12.
- [20]A.L. Robertson, J. Horne, A.M. Ellisdon, B. Thomas, M.J. Scanlon, and S.P. Bottomley, The structural impact of a polyglutamine tract is location-dependent. *Biophys J* 95 (2008) 5922-30.
- [21]A. Bacolla, and R.D. Wells, Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem* 279 (2004) 47411-4.
- [22]C.E. Pearson, M. Tam, Y.H. Wang, S.E. Montgomery, A.C. Dar, J.D. Cleary, and K. Nichol, Slipped-strand DNAs formed by long (CAG)<sup>n</sup>(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Res* 30 (2002) 4534-47.

---

**Figure Legends**

**Figure 1:** There are two reactions involved in the amplification of consecutive CAG repeats. Reaction 1 involves the insertion of a Type IIS restriction site (*BseR1*) at a specific location 7bp upstream from the beginning of the CAG<sub>n</sub> (Qn) repeat. The vector is then digested with *BseR1* and *Nde1*, for which the cleavage site is located 5' to the beginning of the SpA coding sequence. Digestion leaves a linearized vector with Q(n-1) repeats, and an antisense TC overhang. Reaction 2 involves the amplification of the SpA-cQn coding region. The reverse amplification primer introduces another Type IIS recognition site (*Bsg1*) 15bp downstream from the Qn repeat. Subsequent digestion of this fragment with *Bsg1* and *Nde1* leaves a fragment comprising SpA-cQ(n-1) repeats, with an AG overhang 3' on the sense strand. The products of Reactions 1 and 2 are ligated, yielding a construct comprising SpA-cQ(2n-1). This product can then be fed back into the beginning of reaction 2. Subsequent products are amplified in a cyclic process, yielding constructs SpA-cQ(2n-1), SpA-cQ(3n-2), SpA-cQ(4n-3), SpA-cQ(5n-4) etc.

**Figure 2:** 2% agarose gel of PCR-amplified SpA-cQn fragments.

Figure 1

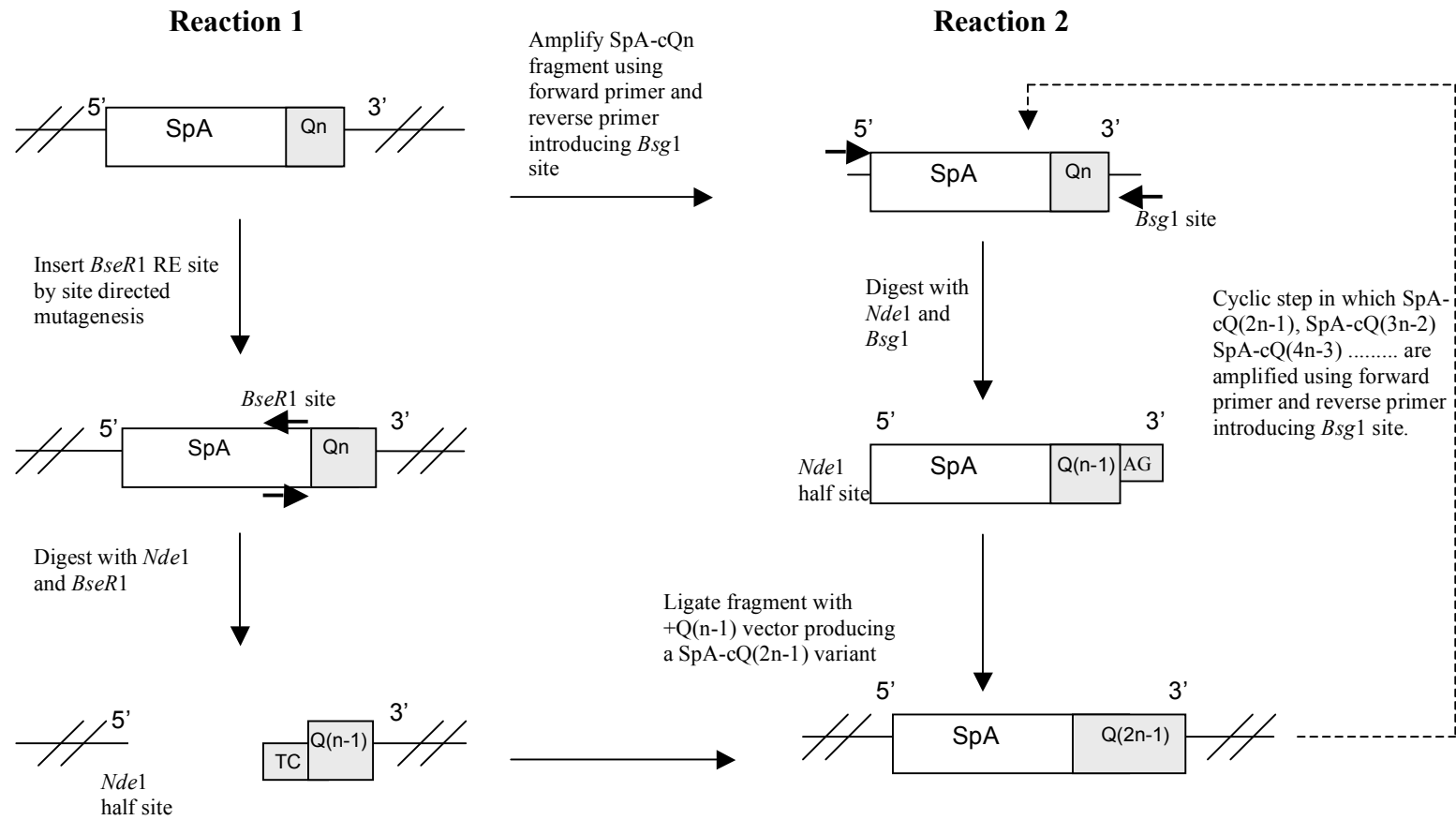
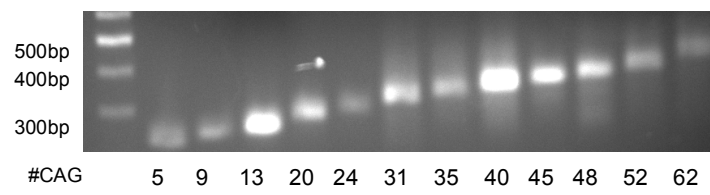


Figure 2





## **Addendum to Chapter 3**

### **Materials and Methods**

## **Materials and Methods**

### **Introduction of *BseR1* restriction site into SpA-cQ5-pET21a**

The DNA sequence encoding the B1 domain of *Staphylococcus aureus* Protein A (SpA) was codon optimized for high-level expression in *E. coli*, and the gene sequence synthesized by TOP gene technologies (Canada). The unique restriction site *BseR1* was introduced using a method based on the QuikChange site-directed mutagenesis protocol (Stratagene) using a pair of complementary mutagenic primers: 5'-TGATGCGCAGGCGAGGAGACCGCGGCAGCA-3' and 5'-TGCTGCCGCGGTCTCCTCGCCTGCGCATCA-3' at a specific site within the SpA sequence so that when cleaved with *BseR1*, a compatible cohesive overhang was left at the 5' end of the CAG repeat. PCR was performed in a 25  $\mu$ L reaction mixture containing 50 ng of template SpA-cQ5 pET21a, 500 nM of each primer, 200  $\mu$ M of each dNTP, 1.45 U of *Pfu* DNA polymerase (Promega) in 1x *Pfu* DNA polymerase buffer (Promega). Thermal cycling was performed as follows: initial denaturation at 95 °C for 1 min; 18 cycles at 95 °C for 1 min, 55 °C for 1 min and 72 °C for 12 min; and final extension at 72 °C for 12 min. After PCR, methylated template strand was digested with 10 U of *Dpn1* (Promega) and incubated at 37 °C for 2 h. 5  $\mu$ l of product was then transformed into 50  $\mu$ l NovaBlue GigaSingles competent cells (Novagen).

### **Vector preparation**

*BseR1* site-inserted SpA-cQ<sub>5</sub> vector was digested with *BseR1* and *Nde1* in 10 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 50 mM NaCl and 1 mM dithiothreitol, pH 7.9, supplemented with 100  $\mu$ g/ml bovine serum albumin, for 2 h at 37 °C. Digested vector was purified using the Promega gel purification kit (according to manufacturer's instructions), and DNA quantified using  $\epsilon_{260}$ =50 ng/ $\mu$ l. This digestion cuts 5' to SpA, leaving an *Nde1* half-site,

and leaves only 4 CAG repeats, plus a TC overhang on the antisense strand, corresponding to the *BseR1* cleavage site (+Q(n-1) vector).

### Fragment preparation

The SpA-cQ<sub>n</sub> fragments were PCR amplified from the pET21a vector using the following primer pair: 5'-CCCGCGAAATTAATACGACTCACTATAG-3' and 5'-ATGCGTGCAGGATCCTCACCCGGGCTGCTGC-3'. The second primer introduces a *Bsg1* recognition sequence. Fragment amplification was performed in a 25 µl reaction volume, containing; 100 ng of template SpA-cQ<sub>n</sub> pET21a, 500 nM of each primer, 200 µM of each NTP, 1.45 U of *Pfu* DNA polymerase (Promega) in 1x *Pfu* DNA polymerase buffer (Promega). The thermal cycler was programmed as follows: initial denaturation at 72 °C for 1 min; 35 cycles at 95 °C for 1 min, 56 °C for 1 min and 72 °C for 1 min, 30 sec; and final extension at 72 °C for 1 min, 30 sec.

The SpA-cQ<sub>n</sub> fragments were digested with *Nde1*(Promega) and *Bsg1*(NEB), in 20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM dithiothreitol, pH 7.9, supplemented with 100 µg/ml bovine serum albumin and S-adenosylmethionine, for 2 h at 37 °C. Digested fragments were then purified using the Promega gel purification kit (according to manufacturer's instructions), and DNA quantitated using  $\epsilon_{260} = 50 \text{ ng}/\mu\text{l}$ . This digestion leaves a fragment with a 5' *Nde1* half-site and an AG overhang at the 3' end corresponding to the *Bsg1* cleavage site.

### Ligations and colony screening

Digested SpA-cQ<sub>n</sub> fragments were ligated with the +Q(n-1) vector using T4 DNA ligase (Promega) for 2 h at room temperature. Ligation products were transformed into

---

competent JM107 cells, and colonies screened by *BseR1* digestion and DNA sequencing.

## **Chapter 4**

# **The Structural Impact of a Polyglutamine tract is Location-dependent**

Published in *Biophysical Journal*, December 2008.

# The Structural Impact of a Polyglutamine Tract Is Location-Dependent

Amy L. Robertson,\* James Horne,<sup>†</sup> Andrew M. Ellisdon,\* Bronwen Thomas,\* Martin J. Scanlon,<sup>†</sup> and Stephen P. Bottomley\*

\*Department of Biochemistry and Molecular Biology, Monash University, Clayton, Australia; and <sup>†</sup>Medicinal Chemistry and Drug Action, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Australia

**ABSTRACT** Polyglutamine (polyQ) expansion leads to protein aggregation and neurodegeneration in Huntington's disease and eight other inherited neurological conditions. Expansion of the polyQ tract beyond a threshold of 37 glutamines leads to the formation of toxic nuclear aggregates. This suggests that polyQ expansion causes a conformational change within the protein, the nature of which is unclear. There is a trend in the disease proteins that the polyQ tract is located external to but not within a structured domain. We have created a model polyQ protein in which the repeat location mimics the flexible environment of the polyQ tract in the disease proteins. Our model protein recapitulates the aggregation features observed with the clinical proteins and allows structural characterization. With the use of NMR spectroscopy and a range of biophysical techniques, we demonstrate that polyQ expansion into the pathological range has no effect on the structure, dynamics, and stability of a domain adjacent to the polyQ tract. To explore the clinical significance of repeat location, we engineered a variant of the model protein with a polyQ tract within the domain, a location that does not mimic physiological context, demonstrating significant destabilization and structural perturbation. These different effects highlight the importance of repeat location. We conclude that protein misfolding within the polyQ tract itself is the driving force behind the key characteristics of polyQ disease, and that structural perturbation of flanking domains is not required.

## INTRODUCTION

Protein aggregation is a hallmark of a number of neurodegenerative diseases, including Alzheimer's disease, Parkinson's disease, prion diseases, and the polyglutamine (polyQ) diseases (1–3). A common feature of all of these diseases is the formation of amyloid-like fibrils with a common  $\beta$ -sheet rich core (4–6). A significant challenge is to determine the structural changes that occur during the formation of fibrillar aggregates, which remain unclear for many of the proteins linked with misfolding and disease.

Nine inherited, late-onset polyQ disorders have been identified, including Huntington's disease (HD) and a number of spinocerebellar ataxias. The disease-linked mutation is the expansion of a polyQ tract beyond a context-specific threshold, which in the case of HD is  $\sim 37$  glutamines. The presence of an expanded polyQ tract leads to the deposition of neuronal inclusions, cell death, and neurodegeneration. The polyQ region is fundamental to disease, and although there are many specific pathological and molecular phenotypes, the ability of proteins with a pathological-length polyQ tract to form fibrillar aggregates is the common feature of these nine diseases (5,7).

Aggregation studies of polyQ peptides and polyQ-containing proteins suggest that the aggregation nucleus is a monomer; therefore, determining the nature of the repeat length-dependent conformational changes within the mono-

meric proteins is central to understanding the pathogenic mechanism. Wetzel and colleagues (4) demonstrated that the driving force for the aggregation of isolated polyQ peptides is a random coil to  $\beta$ -sheet conformational conversion. Peptide studies have provided key information regarding the isolated polyQ tract; however, there is evidence to suggest that protein context has an important role in polyQ-derived pathogenesis. Previous studies have shown that the composition of the polyQ flanking domains affects the solubility and toxicity of pathological-length polyQ proteins (8–11). Recent findings indicating that proteins with expanded polyQ tracts have mechanistic perturbations, and that aggregates may be comprised of proteolytically cleaved fragments of polyQ-containing proteins suggest that structural and/or dynamic changes outside of the polyQ domain may be a common consequence of repeat expansion (12–15). Indeed, Perutz (16) hypothesized that proteins with pathological-length polyQ tracts would be destabilized and hence the structure and dynamics of the flanking domains would be altered. The native-state destabilization hypothesis has been investigated with the use of ataxin-3 (17,18), a CRABP1-Htt exon1 fusion (19), and a myoglobin-polyQ model protein (20). These studies were hindered by several factors. The N-terminal Josephin domain of ataxin-3 aggregates independently of polyQ expansion (21,22). Htt exon 1 contains a C-terminal oligoproline motif, and there is evidence that this modulates the conformational properties of the polyQ repeat (23). In the myoglobin-polyQ study, Tanaka et al. (20) inserted the polyQ tract in a loop within the myoglobin domain; however, within the disease-associated proteins the tracts are found either N- or C-terminal to other protein domains (24). Although it is

Submitted May 28, 2008, and accepted for publication September 4, 2008.

Address reprint requests to Stephen P. Bottomley, Dept. of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia. Tel.: 61-3-99053703; Fax: 61-3-99054699; E-mail: steve.bottomley@med.monash.edu.au.

Editor: Doug Barrick.

© 2008 by the Biophysical Society  
0006-3495/08/12/5922/09 \$2.00

doi: 10.1529/biophysj.108.138487

clear that conformational changes in the polyQ tract due to expansion lead to aggregation, one unresolved question is, Does polyQ expansion lead to a change in the structure and dynamics of surrounding domains? This is an important issue in terms of therapeutic approaches because it would be far easier to target and stabilize the folded domains surrounding the polyQ tract than the polyQ region itself.

Here we present the first, to our knowledge, structural characterization of a domain fused to polyQ tracts of different lengths. We report that the stability, structure, and dynamics of the flanking domain do not change when the polyQ tract expands into the pathological range, and hence conclude that it is the intrinsic propensity of the polyQ domain to adopt a  $\beta$ -sheet structure that drives aggregation, a mechanism consistent with polyQ peptide aggregation. Furthermore, we show significant structural perturbation when the polyQ tract is engineered within a domain, which suggests that repeat tract location determines the misfolding pathway, and therefore the choice of location is essential in the design of model proteins. These location-dependent effects point toward the clinical significance of discrete repeat domains.

## MATERIALS AND METHODS

### Protein expression and purification

The amino acid sequence of the B domain of *Staphylococcus aureus* Protein A was codon-optimized and the gene was synthesized by TOP Gene Technologies (Quebec, Canada) and inserted into our pLIC-His construct for expression (25). A construct with 5 C-terminal CAG repeats was also synthesized (SpA-cQ5). To create the intradomain polyQ mutants, a complementary primer pair was designed to introduce five CAG repeats between proline 39 and serine 40 of SpA using site-directed mutagenesis. Based on the method of Peters and Ross (26), a method to incrementally amplify the number of CAG repeats was devised that allowed the construction of C-terminal fusion proteins with polyQ repeat lengths ranging from 5 to 52 glutamines and intradomain polyQ variants with 9 and 25 glutamines inserted between helices 2 and 3 of SpA.

Unlabeled SpA-polyQ fusion proteins were expressed in BL21(DE3) cells by autoinduction (27). Labeled proteins for NMR uniformly enriched with either  $^{15}\text{N}$  only or  $^{15}\text{N}/^{13}\text{C}$  were expressed according to the protocol of Marley et al. (28). Proteins were purified with the use of a two-step chromatography procedure and the Akta Express protein purification system. A nickel chelating chromatographic step was performed with the use of a 1 mL HisTrap (GE Healthcare, Uppsala, Sweden) column according to the manufacturer's instructions. Proteins were further purified by gel filtration chromatography with a Superdex 75 16/60 or Superdex 200 16/60 column (Pharmacia). Proteins were eluted and stored in the gel filtration buffer (20 mM sodium acetate (pH 5.0), 100 mM NaCl, 10% (w/v) glycerol). Protein concentration was determined by measuring the  $\text{Abs}_{280}$  of protein samples unfolded with guanidine thiocyanate. The purity of proteins was analyzed by means of 12% Tris-tricine sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

### Size exclusion chromatography

Size exclusion chromatography (SEC) of SpA-polyQ proteins was performed on an Akta-FPLC equipped with a Superdex 75 10/300 GL gel filtration column (Pharmacia). The column was precalibrated with proteins of known molecular mass. SpA-polyQ samples at a concentration of 50  $\mu\text{M}$  were applied to the column, and UV absorption at 280 nm was monitored as

the proteins were eluted in 20 mM sodium acetate (pH 5.0), 100 mM NaCl, 2 mM phenylmethylsulfonyl fluoride (PMSF), 10% (w/v) glycerol at a constant flow rate of 0.5 mL/min.

### Circular dichroism

Circular dichroism (CD) spectra were measured on a Jasco-810 spectropolarimeter at 25°C by means of a cuvette with a pathlength of 0.01 cm and protein concentration of 30  $\mu\text{M}$ . Spectra were recorded from 190 to 260 nm with a scan speed of 50 nm/min. The spectra were analyzed by deconvolution with the K2D algorithm as provided by the DICHROWEB online facility (29).

### Protein aggregation time-course assays

SpA-polyQ proteins were incubated at a concentration of 20  $\mu\text{M}$  in 20 mM sodium acetate (pH 5.0), 100 mM NaCl, 2 mM PMSF, 10% (w/v) glycerol. Reactions were performed at 37°C without shaking and in airtight containers to eliminate evaporation.

To monitor aggregation with the use of ThioflavinT (ThT), reactions were performed in 384-well, clear-bottom plates in the presence of 20  $\mu\text{M}$  ThT, and fluorescence emission intensity at 490 nm was recorded ( $\lambda_{\text{ex}} = 450$ ) and measured on a BMG Laboratories FLUOstar Optima fluorescence plate reader. Protein concentrations, buffering solutions, and additives were the same as described above. Reactions were carried out at 37°C, without shaking, and the top of the plate was sealed to prevent evaporation.

PolyQ aggregation inhibition assays were performed with the QBP1 peptide. QBP1 peptide was purchased from AusPep with the same sequence as described previously (30). The peptide was dissolved in 50%  $\text{Me}_2\text{SO}/50\%$  (20 mM sodium acetate (pH 5.0), 100 mM NaCl, 10% (w/v) glycerol) and incubated with SpA-polyQ aggregation assays (as above) at a final concentration of 120  $\mu\text{M}$ . Control assays with identical final  $\text{Me}_2\text{SO}$  concentrations were performed and showed no significant effect of  $\text{Me}_2\text{SO}$  addition.

### Membrane filter trap assay

This assay was carried out with minor modifications to the previously described protocol (31). A 10  $\mu\text{L}$  aliquot of protein 20  $\mu\text{M}$  was removed from the aggregation time-course incubations and diluted at a ratio of 1:1 in a solution of 4% SDS (w/v) and 100 mM dithiothreitol. The samples were heated for 5 min at 100°C before being diluted to 150  $\mu\text{L}$  with 2% (w/v) SDS. The samples were filtered through a cellulose acetate membrane (Schleicher and Schuell, 0.2  $\mu\text{m}$ ) using a Bio-Rad Bio-Dot SF microfiltration unit. After application, the membrane was washed twice by filtering through 150  $\mu\text{L}$  of 0.1% (w/v) SDS. Immunoblot analysis was performed using a primary anti-His antibody (Serotec).

### Transmission electron microscopy

Transmission electron microscopy (TEM) images were obtained using an Hitachi H7500 transmission electron microscope. The acceleration voltage was 80 kV. The samples were adsorbed onto a carbon-coated grid and stained with 1% (w/v) uranyl acetate.

### Thermal and guanidine denaturation

Protein at a concentration of 20  $\mu\text{M}$  was denatured as the temperature was increased from 40°C to 90°C at a rate of 1°C/min. Samples in the presence or absence of 1 M GuHCl were in a thermostated cuvette with a pathlength of 0.1 cm, and the CD signal at 222 nm was measured.

For chemical denaturation experiments, guanidine hydrochloride (GuHCl) was prepared in 20 mM sodium acetate (pH 5.0), 100 mM NaCl, 10% (w/v) glycerol. Equilibrium unfolding titrations were set up with GuHCl ranging from 0 to 6 M, with a protein concentration of 20  $\mu\text{M}$ . Proteins were incubated

in GuHCl solutions at room temperature for 30 min, and the CD signal at 222 nm ( $CD_{222}$ ) was recorded. To determine reversibility, proteins unfolded in 6 M GuHCl were titrated back into Buffer A and  $CD_{222}$  was recorded.

## NMR experiments and SpA-cQ5 structure determination

Samples for NMR analysis were prepared to a concentration of 1 mM and 150  $\mu$ M of SpA-cQ5 and SpA-cQ52, respectively, in 20 mM sodium acetate (pH 5.0), 100 mM NaCl, 10% (w/v) glycerol in 90%  $H_2O$ : 10%  $^2H_2O$ . All experiments were recorded on a Varian Unity 600 MHz spectrometer equipped with a triple resonance cold probe with actively shielded gradients at a temperature of 298 K. For the comparison of SpA-cQ5 and SpA-cQ52, 10% (w/v) glycerol was added to both samples to improve the SpA-cQ52 sample stability and to confirm that there was no significant effect on the SpA-cQ5 spectra.

Standard triple resonance experiments were recorded to obtain assignments for SpA-cQ5 (32). Two- and three-dimensional nuclear Overhauser enhancement spectroscopy (NOESY) experiments were recorded for measurement of interproton distances (33).

Data were processed with the use of NMRPipe (34) and analyzed with the software SPARKY (Goddard and Kneller, University of California, San Francisco, CA). Initial structures were calculated in Cyana-2.1 by means of automated NOE assignment (35). Backbone  $\phi/\psi$  angle restraints were calculated based on chemical shifts for  $C\alpha$ ,  $C\beta$ ,  $H\alpha$ ,  $HN$ , and  $C'$  nuclei with the software TALOS (36). The final structures were refined with Xplor-NIH (37). Hydrogen-bond restraints were added for amide protons with measured slow exchange in CLEANEX-PM (38) experiments and for which a hydrogen bond acceptor could be identified clearly in the initial structure ensembles. The ensemble of 29 structures had no NOE violations  $> 0.3$  Å or dihedral violations  $> 5^\circ$  and good overall geometry as determined by Whatcheck and Procheck-NMR, and was deposited in the Protein Data Bank (PDB id: 2JWD).

## RESULTS

### PolyQ proteins with repeat lengths of $\geq 35$ form fibrillar aggregates

We constructed a family of proteins containing variable-length polyQ regions attached to the C terminus of the Protein A B domain (Y15W mutant) from *Staphylococcus aureus* (SpA) (39–41). The length of the polyQ tract within the polyQ protein increases incrementally, from five to 52 glutamines (Fig. 1 *a*). Size-exclusion chromatography demonstrated that all of the proteins were entirely monomeric (Fig. 1 *b*). The elution volume of the protein increased linearly with the number of glutamines, indicating that there is no thermodynamically favorable conformational change within the monomeric proteins with increasing polyQ length. This observation is consistent with previous data indicating that, regardless of repeat length, polyQ tracts exist in a random coil in solution (42). This was further confirmed by far-UV CD analyses of the polyQ proteins (Fig. 1 *c*). The presence of the polyQ tract causes little change in the SpA spectra, with spectral deconvolution, indicating that the polyQ is random coil in structure, with a proportional decrease in  $\alpha$ -helical content from 79% in SpA-Q0 to 48% in SpA-cQ52, and a concomitant increase in fraction of random coil from 20% in SpA-Q0 to 40% in SpA-cQ52.

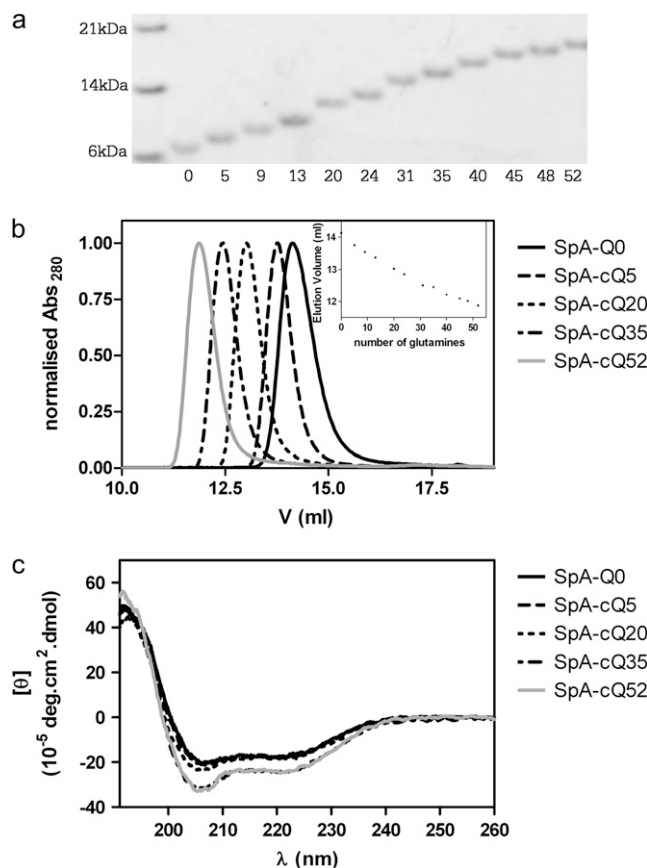


FIGURE 1 SpA-polyQ proteins are monomeric and the polyQ region is unstructured. (a) Purified SpA-polyQ fusion proteins were analyzed by 12% (w/v) SDS-PAGE. (b) SpA-polyQ fusion proteins were analyzed by SEC using a Superdex 75 column. The elution profiles of SpA-Q0, SpA-cQ5, SpA-cQ20, SpA-cQ35, and SpA-cQ52 are shown. (Inset) The elution volumes of all 12 proteins are plotted against the number of glutamines on the inset graph, showing a linear relationship. (c) Far-UV CD spectra of SpA-polyQ proteins. The spectra of SpA-Q0, SpA-cQ5, SpA-cQ20, SpA-cQ35, and SpA-cQ52 are shown. There is increasing random coil content with expansion of the polyQ region.

Only SpA-polyQ proteins with polyQ tracts of 35 or greater formed ThT-reactive aggregates with a rate dependent on the repeat length (Fig. 2 *a*). The duration of the lag phase decreased exponentially as a function of polyQ length as the number of glutamines increased from 35 to 52 repeats. This is indicated by the change in the midpoint of the ThT curves, going from  $\sim 270$  h for SpA-cQ35 and decreasing exponentially to plateau at  $\sim 10$ – $15$  h for proteins with 48–52 glutamines (Fig. 2 *a*, inset). Analysis of the kinetics of aggregation also demonstrated both time and concentration dependence, consistent with the nucleation-polymerization aggregation model proposed for polyQ aggregation (43) (data not shown). In addition, the aggregates were found to be resistant to solubilization by SDS and their formation could be inhibited by the QBP1 peptide (30,31) (Fig. 2 *b*). Morphological analysis of each protein revealed that regardless of polyQ length, all form similar amyloid-like fibrillar struc-



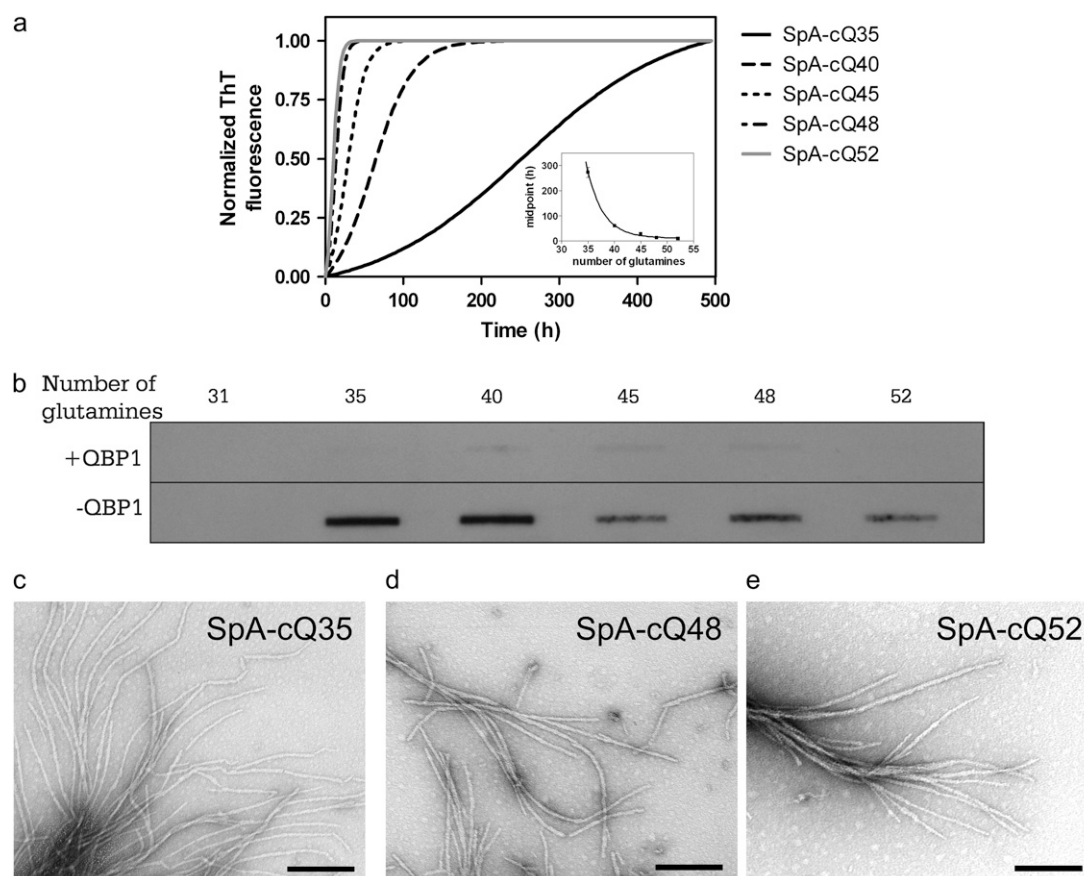


FIGURE 2 Expanded PolyQ proteins form typical fibrils. (a) ThT analysis of aggregation formation for proteins with repeat lengths  $\geq 35$ . Proteins with repeat lengths below 35 did not aggregate over the 500 h timespan analyzed. Curves shown are single exponential fits of data sets. (Inset) The midpoints of the ThT traces plotted as a function of repeat length. (b) Analysis of the SDS stability of endpoint aggregates in the presence (top panel) and absence (bottom panel) of QBP1. (c–e) TEM analysis SpA-cQ35 (c), SpA-cQ48 (d), and SpA-cQ52 (e). Scale bars are 200 nm.

tures, which are unbranched and range from 7 to 15 nm in diameter (Fig. 2, c–e). There was no observation of fibrils for SpA-cQ31 or variants with shorter repeat lengths. Taken together, the aggregation data suggest that the expanded SpA-polyQ proteins form amyloid-like fibrils with a disease threshold, structure, and kinetic mechanism similar to those of the disease-causing proteins.

### Thermodynamic stability is not linked to polyQ expansion

The increased aggregation propensity of SpA-polyQ variants with  $\geq 35$  glutamines, combined with evidence that the composition of flanking regions influences aggregation and toxicity, suggests that mutant polyQ expansions may affect the conformation of the surrounding domains. To investigate this, we performed thermal and guanidine (GuHCl) equilibrium un/folding experiments on the 12 SpA-polyQ proteins. Of interest, we found that all 12 proteins displayed equivalent thermal and GuHCl-induced unfolding profiles, with midpoints for the unfolding transitions of  $\sim 69^\circ\text{C}$  and 4 M, respectively (Fig. 3, a and b, Table 1). Guanidine-induced

unfolding transitions were found to be two-state and fully reversible.  $\Delta G^\circ$  values were similar for all 12 proteins (0–52 Qs), ranging from 6 to 7 kcal/mol, indicating that polyQ expansion had no effect on protein stability. Many proteins, under conditions that cause partial unfolding (such as low denaturant concentrations in which hydrophobic interactions are disrupted), are prone to spontaneous aggregation. To further confirm that SpA-polyQ protein samples in GuHCl-induced folding titrations remained monomeric, ThT and SEC analyses of samples in the presence of 0–6 M GuHCl were performed, and the results indicated that there were no aggregates present (data not shown). The fact that the presence of guanidine does not induce aggregation of proteins with an expanded polyQ tract is consistent with a model in which aggregation is entirely driven by the polyQ region.

### The structure and dynamics of the flanking domain are unaffected by polyQ expansion

The un/folding data suggest that expansion of the polyQ region does not affect the conformation of SpA; however, this is a global stability measure and offers no insight at the amino

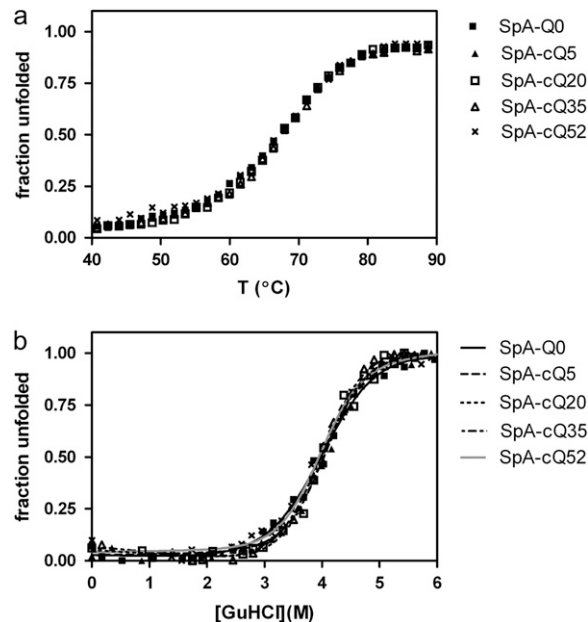


FIGURE 3 PolyQ expansion does not affect thermal or GuHCl-induced denaturation profiles. (a) All polyQ proteins showed similar thermal unfolding profiles. The change in CD signal at 222 nm was monitored as proteins were heated from 40°C to 90°C in the presence of 1 M GuHCl. (b) All polyQ proteins showed similar unfolding profiles when unfolded with GuHCl. SpA-Q0, SpA-cQ5, SpA-cQ20, SpA-cQ35, and SpA-cQ52 are shown as representative traces and thermal and chemical denaturation profiles of all proteins overlaid (Table 1).

acid level. We utilized NMR to specifically assess the impact of polyQ expansion on the structure of the SpA domain to determine whether there might be local conformational changes induced by polyQ expansion. The NMR experiments also provided crucial insight into the dynamic properties of the SpA domain in the different proteins. Such

TABLE 1 Equilibrium stability of C-terminal SpA-polyQ proteins

|          | Tm (°C)*   | Dm (M) <sup>†</sup> | ΔG° (kcal/mol) <sup>‡</sup> |
|----------|------------|---------------------|-----------------------------|
| SpA-Q0   | 69.0 ± 0.4 | 4.1                 | 6.5 ± 0.2                   |
| SpA-cQ5  | 69.3 ± 0.5 | 4.2                 | 6.8 ± 0.3                   |
| SpA-cQ9  | 69.3 ± 0.3 | 4.1                 | 6.6 ± 0.3                   |
| SpA-cQ13 | 68.8 ± 0.3 | 4.1                 | 6.5 ± 0.5                   |
| SpA-cQ20 | 68.7 ± 0.4 | 4.1                 | 6.6 ± 0.2                   |
| SpA-cQ24 | 69.4 ± 0.3 | 4.1                 | 6.5 ± 0.5                   |
| SpA-cQ31 | 70.0 ± 0.4 | 3.9                 | 6.2 ± 0.3                   |
| SpA-cQ35 | 69.4 ± 0.5 | 4.2                 | 6.7 ± 0.2                   |
| SpA-cQ40 | 68.7 ± 0.5 | 4.0                 | 6.4 ± 0.2                   |
| SpA-cQ45 | 68.8 ± 0.4 | 4.1                 | 6.5 ± 0.5                   |
| SpA-cQ48 | 69.4 ± 0.4 | 4.1                 | 6.5 ± 0.5                   |
| SpA-cQ52 | 69.3 ± 0.5 | 4.2                 | 6.7 ± 0.2                   |

\*Midpoint of the thermal denaturation curves determined by measuring the change in CD signal at 222 nm in the presence of 1 M GuHCl.

<sup>†</sup>Midpoints of the GuHCl-denaturation curves determined using the equation for two-state unfolding.

<sup>‡</sup>ΔG° was determined by analysis of the GuHCl-denaturation curves, with a shared equilibrium *m* value of 1.6 kcal/mol/M.

information is invaluable for determining whether a pathological-length polyQ tract alters the conformational heterogeneity of the native ensemble. We determined the NMR structure of SpA-cQ5, which remained essentially unchanged from that of the previously reported NMR structure for the SpA domain (40) (Fig. 4, *a* and *b*; Table 2). To identify structural changes caused by polyQ expansion into the pathological range, we performed <sup>1</sup>H-<sup>15</sup>N-HSQC experi-

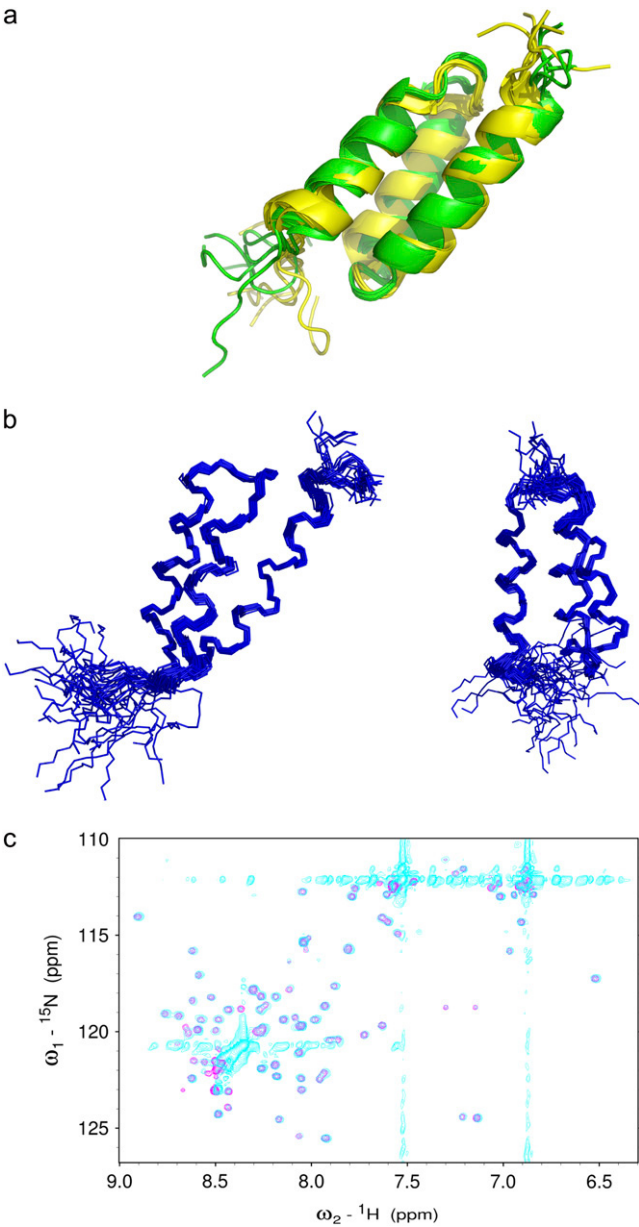


FIGURE 4 Structure of the flanking domain. (a) Backbone atom (N, Cα, C') trace of the ensemble of 29 structures of SpA-cQ5 and 90° rotation around the y axis. (b) Backbone atom superposition of the five structures with the lowest energy from our ensemble of SpA-cQ5 (green) with the first five structures from the published 1SS1.pdb NMR ensemble for SpA (yellow). (c) Overlays of <sup>1</sup>H-<sup>15</sup>N-HSQC spectra for SpA-cQ5 (magenta) and SpA-cQ52 (cyan) recorded at 600 MHz and 25°C.

**TABLE 2** NMR and refinement statistics for 29 SpA-cQ5 model structures

|  |                      |
|--|----------------------|
| NMR distance and dihedral constraints                |                      |
| Distance constraints                                 |                      |
| Total NOE  |                      |
| Interresidue   | 657                  |
| Short-range ( $ i-j  = 1$ )                          | 237                  |
| Medium-range ( $ i-j  < 4$ )                         | 253                  |
| Long-range ( $ i-j  > 3$ )                           | 97                   |
| Total dihedral angle restraints                      |                      |
| $\phi$   | 37                   |
| $\psi$   | 38                   |
| H-bond distance restraints                           | 14                   |
| Structure statistics                                 |                      |
| Violations (mean and SD)                             |                      |
| Distance constraints (Å)                             | $0.0325 \pm 0.0041$  |
| Dihedral angle constraints (°)                       | $0.3187 \pm 0.1571$  |
| Deviations from idealized geometry                   |                      |
| Bond lengths (Å)                                     | $0.0038 \pm 0.00018$ |
| Bond angles (°)                                      | $0.7338 \pm 0.0265$  |
| Impropers (°)  | $0.6903 \pm 0.0354$  |
| Lennard-Jones energy (kcal mol <sup>-1</sup> )*      | $-258.91 \pm 7.82$   |
| Root mean-square deviation to the mean structure (Å) |                      |
| Backbone   | $0.387 \pm 0.124$    |
| Heavy  | $0.878 \pm 0.158$    |

Pairwise root mean-square deviation was calculated among 29 refined structures.

\*Lennard-Jones energy values from CNS using protein par parameters from the Xplor-NIH distribution.

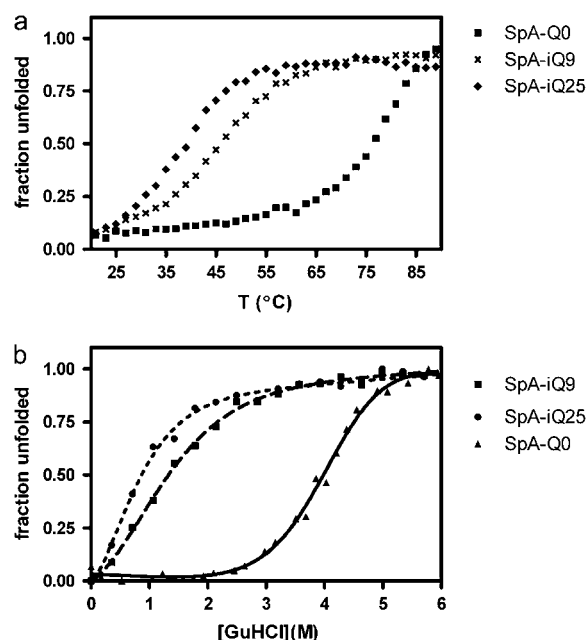
ments on SpA-cQ52. A comparison of the <sup>1</sup>H and <sup>15</sup>N chemical shifts for residues in the SpA domain of SpA-cQ5 and SpA-cQ52 revealed no significant chemical-shift perturbations. This suggests that the structure of SpA is not perturbed even when fused to 52 glutamines, as demonstrated by the overlaid <sup>1</sup>H-<sup>15</sup>N-HSQC spectra (Fig. 4 c). The redundant glutamine chemical shifts in the SpA-cQ52 spectra suggest that the polyQ tail exists in a random coil conformation, consistent with our far-UV CD data. To retard aggregation of the SpA-cQ52 sample, an ~7-fold lower protein concentration was used as compared to the SpA-cQ5 sample; therefore, a direct comparison of the peak intensities between the spectra is uninformative. Analysis of the peaks across the individual spectra, however, reveals that in each data set the relative intensities of the cross peaks remain similar. That is, none of the cross peaks in the <sup>1</sup>H-<sup>15</sup>N HSQC data of SpA-cQ52 are either selectively broadened or changed in intensity compared to the spectrum of SpA-cQ5. These data suggest that polyQ expansion into the pathological range does not result in perturbation of the structure of the flanking domain, and that the SpA-cQ5 and SpA-cQ52 proteins are similar in terms of their dynamics.

### A polyQ tract disrupts tertiary interactions when located within the domain

To explore the significance of repeat location, we engineered a variant of the SpA-polyQ model protein with a polyQ tract

within a flexible loop of SpA to observe the effects of an intradomain polyQ repeat on host domain stability and structure (in a nonphysiological location). We engineered insertion mutants with the polyQ tract placed in the flexible loop between helices 2 and 3 of SpA, the insertion site indicated by the arrow in Fig. 6 c. Poor protein expression limited analyses to variants with repeat lengths below the pathological range, and as a result we did not observe aggregation of these variants; however, we did observe destabilization of variants with Q9 and Q25 inserts (denoted SpA-iQ9 and SpA-iQ25, respectively). As compared to wild-type SpA (SpA-Q0), the SpA-iQ9 variant was significantly destabilized by ~4.7 kcal/mol, and SpA-iQ25 was further destabilized (Fig. 5 a, Table 3). Decreased thermal stability was also observed with shifts in the thermal denaturation midpoints with increasing repeat length (Fig. 5 b, Table 3).

We assessed the effect of the intradomain Q9 and Q25 inserts on the structure of SpA by CD and NMR. CD analyses demonstrated that the SpA-iQ9 and SpA-iQ25 variants retain their  $\alpha$ -helical secondary structure (Fig. 6 a). The polyQ insertions, however, affect the tertiary structure of SpA, as demonstrated by the high-field methyl regions in the 1D <sup>1</sup>H NMR spectra of SpA, SpA-cQ5, and SpA-iQ9 (Fig. 6 b). The spectrum of SpA contains several resonances for methyl groups with chemical shifts of <0.5 ppm, as would be expected in a folded protein domain. The highest field resonances are observed for Ile<sup>17</sup>, Leu<sup>18</sup>, Leu<sup>20</sup>, and Ala<sup>49</sup>. The location of these methyl groups in the tertiary structure of SpA is highlighted in Fig. 6 c. The methyls



**FIGURE 5** Intradomain polyQ tract affects the stability of SpA. (a) Thermal denaturation of SpA-Q0, SpA-iQ9, and SpA-iQ25. The change in CD signal was monitored as proteins were unfolded from 20°C to 90°C. (b) GuHCl unfolding of SpA-Q0, SpA-iQ9, and SpA-iQ25. SpA-iQ9 and SpA-iQ25 unfold in lower concentrations of GuHCl relative to SpA-Q0.

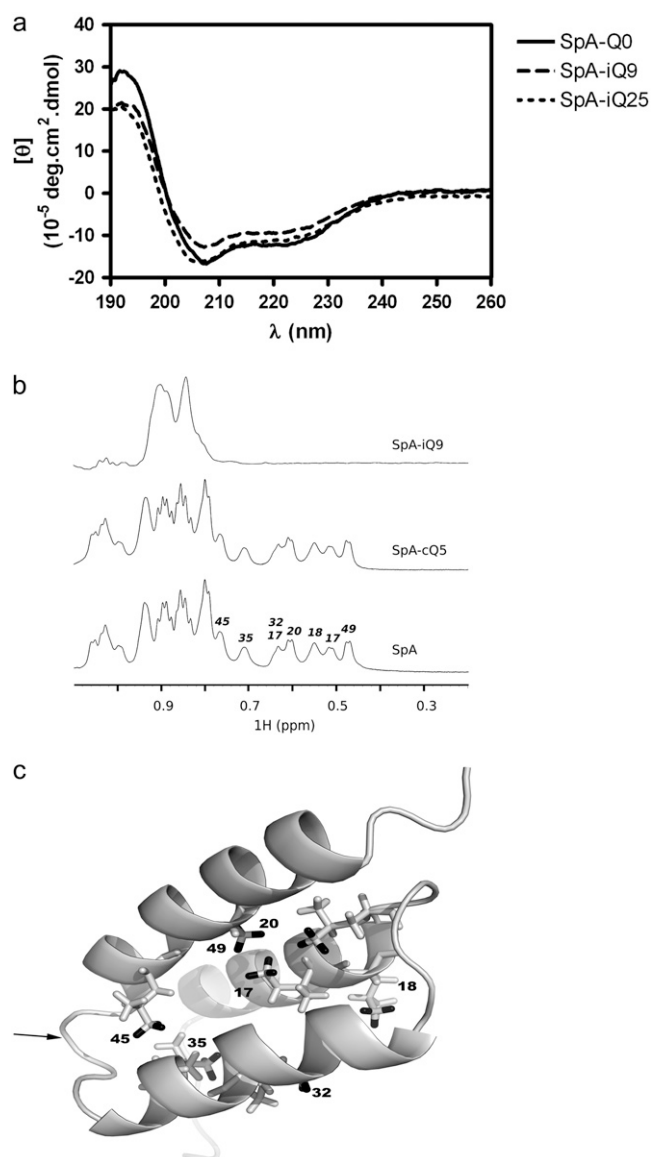


FIGURE 6 Intradomain polyQ repeat structurally perturbs SpA. (a) CD spectra of SpA-Q0, SpA-iQ9, and SpA-iQ25 showing that  $\alpha$ -helical SpA secondary structure is retained in SpA-iQ9 and SpA-iQ25 variants. (b) The high-field region of the  $^1\text{H}$ -NMR spectra of SpA-Q0, SpA-cQ5, and SpA-iQ9, respectively. (c) Cartoon representation of the structure of SpA demonstrating the location of the methyl groups that resonate at the highest field in the  $^1\text{H}$ -NMR spectrum.

of Ile<sup>17</sup>, Leu<sup>20</sup>, and Ala<sup>49</sup> are clustered in the interior of the SpA fold. These resonances are unperturbed upon addition of glutamine extensions at the C terminus of SpA. However, upon insertion of glutamine residues within the SpA domain, there are notable changes in the spectrum. The high-field methyl resonances are lost and collapse to the 0.8–1.0 ppm region. The change in the  $^1\text{H}$  NMR spectrum of SpA-iQ9 is consistent with a destabilization of the tertiary structure of the protein upon insertion of the glutamine repeats. Similar perturbations were observed in the spectra of SpA-iQ25 (data not shown).

TABLE 3 Equilibrium stability of intradomain SpA-polyQ proteins

|          | T <sub>m</sub> (°C)* | D <sub>m</sub> (M) <sup>†</sup> | $\Delta G^\circ$ (kcal/mol) <sup>‡</sup> |
|----------|----------------------|---------------------------------|--|
| SpA-Q0   | — <sup>§</sup>       | 4.1                             | $6.5 \pm 0.2$                            |
| SpA-iQ9  | $46.6 \pm 1$         | 1.5                             | $1.8 \pm 0.3$                            |
| SpA-iQ25 | $38.9 \pm 1$         | — <sup>¶</sup>                  | — <sup>¶</sup>                           |

\*Midpoints of thermal denaturation were determined by measuring the change in CD signal at 222 nm.

<sup>†</sup>Midpoints of the GuHCl-denaturation curves determined using the equation for two-state unfolding.

<sup>‡</sup> $\Delta G^\circ$  was determined by analysis of the GuHCl-denaturation curves, with a shared equilibrium  $m$  value of 1.6 kcal/mol/M.

<sup>§</sup>T<sub>m</sub> for SpA-Q0 could not be determined due to lack of post-transition baseline.

<sup>¶</sup>D<sub>m</sub> and  $\Delta G^\circ$  could not be determined for SpA-iQ25 as data could not be fit to two-state unfolding equation due to absence of pretransition baseline.

## DISCUSSION

In all polyQ-related diseases, the length of the polyQ expansion defines both the age of onset and disease duration. The specific features of each disease, however, suggest that protein context plays a key role in modulating the formation of neuronal aggregates, and this is supported by *in vitro* analyses (8,11). In the disease proteins, the polyQ tract initially exists as a flexible solvent-exposed domain (24). We have replicated this environment with polyQ regions fused to the C terminus of an  $\alpha$ -helical model domain, SpA. The relationship between polyQ repeat length and aggregation seen here recapitulates the exponential association between repeat length and age of HD onset, and the *in vitro* aggregation rates of Htt exon1 proteins (5,7,44). Further, only SpA-polyQ proteins with  $\geq 35$  glutamines formed SDS-stable and morphologically observable fibrils. Our system is also supported by the observation that a repeat length of at least 35 glutamines is required for aggregation initiation, similar to the threshold seen in disease. Taken together, these observations suggest that the SpA domain chosen as a model in this study behaves similarly to the flanking domains in the disease proteins, validating the relevance of our stability and structural data.

Here we examined the hypothesis proposed by Perutz (16), that polyQ expansion causes destabilization and conformational rearrangements of flanking domains. We find that, regardless of repeat length, the C-terminal SpA-polyQ fusion proteins have equivalent thermodynamic stability. These unfolding experiments provide information on the global free energy change, with contributions from the polyQ region and flanking SpA domain. It could be anticipated that the local stability of the polyQ region may change with increasing repeat length. The increased proportion of unstructured conformation, as identified by our CD analyses, would increase the main chain entropy, and unless this is offset by an equivalent increase in solvent entropy, this effect would be reflected by a lowered free energy of the native state. These changes could therefore mask the detection of small changes in free energy contributed by structural alterations of the flanking domain.

We used NMR to more specifically determine the effect of polyQ expansion on the structure of the SpA domain. The  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra for SpA-cQ5 and SpA-cQ52 demonstrate that the SpA domain, even when fused to a pathological-length polyQ region, remains structurally equivalent to SpA alone. Further, the comparable peak intensities across the SpA-cQ5 and SpA-cQ52 spectra suggest that there is a similar degree of conformational heterogeneity for both proteins, supporting the idea of an equivalent equilibrium between the native and unfolded states.

An analysis of all repeat-containing proteins by Faux et al. (24) showed that 73% of all homopeptide repeats are located N' or C' to a structured domain. Of the nine disease-related polyQ proteins, at least seven have polyQ tracts located in flexible, unstructured regions N- or C-terminal to a folded domain. Given the preference of poly amino acid tracts to be located external to a folded domain, we assessed the significance of this in a polyQ protein by analyzing a protein with an intradomain polyQ tract. We observed significant destabilization of the protein and perturbation of the tertiary structure even with a short polyQ tract (SpA-iQ9). The effect of the intradomain polyQ insertion in this case is similar to other model polyQ proteins in which a polyQ tract was inserted within a domain (20,45).

The location-dependent effects of the polyQ tract on the native fold suggest that an expanded repeat tract can induce alternative misfolding pathways. Our data for the C-terminal polyQ proteins are consistent with, and build upon, the model for polyQ peptides in which conformational changes intrinsic to the polyQ segment are sufficient for aggregation to proceed. This contrasts with the significant perturbation we saw when the polyQ tract was placed within a flexible loop of SpA, suggesting that intrinsic polyQ conversion and native-state destabilization can be coincident misfolding pathways. In light of the finding that polyQ expansion within Htt exon1 does destabilize the adjacently fused  $\beta$ -barrel protein CRABP1, our data add complexity to the picture of polyQ-induced misfolding. SpA and CRABP1 differ in native-state stability and topological arrangements, offering alternative models for the conformational properties of the polyQ flanking domains. The ultimate goal, which currently is impeded by the size and solubility of the polyQ disease proteins, is to determine the properties of their flanking domains. This would be of significant interest because the varied results of model polyQ proteins suggest that location and context are key determinants of the perturbing influence of polyQ expansion.

Our NMR data clearly show that polyQ expansion does *not* affect the stability of a folded flanking domain when fused at the C terminus, a result in contrast to another model protein (19). To our knowledge, this is the first study to provide direct structural information about a flanking domain attached to a polyQ tract. Our result does not preclude a perturbing influence of polyQ expansion on surrounding polypeptide regions, but rather suggests that this relationship is complex,

and this is supported by our data showing compromised stability and structure of an intradomain polyQ chimera. The result presented here differs from the classical misfolding model proposed for amyloid formation, in which destabilization of the native state leads to fibril formation in a number of diseases (46–48). Alteration of native-state stability, however, is not a necessary mechanism underlying amyloid formation for a number of other disease-causing proteins, including the prion protein, in which disease-associated mutations do not necessarily alter thermodynamic properties (49,50). In the greater context of amyloid-deposition diseases, the result presented here further highlights the variable pathways along which proteins can aggregate to form uniform amyloid-like aggregates.

S.P.B. is a senior research fellow of the National Health and Medical Research Council (NHMRC), and the work was supported by both the Australian Research Council and NHMRC of Australia.

## REFERENCES

1. Dobson, C. M. 2004. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* 15:3–16.
2. Ross, C. A., and M. A. Poirier. 2005. Opinion: what is the role of protein aggregation in neurodegeneration? *Nat. Rev. Mol. Cell Biol.* 6:891–898.
3. Soto, C. 2003. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.* 4:49–60.
4. Chen, S., F. A. Ferrone, and R. Wetzel. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. USA* 99:11884–11889.
5. Scherzinger, E., A. Sittler, K. Schweiger, V. Heiser, R. Lurz, R. Hasenbank, G. P. Bates, H. Lehrach, and E. E. Wanker. 1999. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc. Natl. Acad. Sci. USA* 96:4604–4609.
6. Yang, W., J. R. Dunlap, R. B. Andrews, and R. Wetzel. 2002. Aggregated polyglutamine peptides delivered to nuclei are toxic to mammalian cells. *Hum. Mol. Genet.* 11:2905–2917.
7. Wanker, E. E. 2000. Protein aggregation and pathogenesis of Huntington's disease: mechanisms and correlations. *Biol. Chem.* 381: 937–942.
8. Duennwald, M. L., S. Jagadish, F. Giorgini, P. J. Muchowski, and S. Lindquist. 2006. A network of protein interactions determines polyglutamine toxicity. *Proc. Natl. Acad. Sci. USA* 103:11051–11056.
9. Duennwald, M. L., S. Jagadish, P. J. Muchowski, and S. Lindquist. 2006. Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc. Natl. Acad. Sci. USA* 103:11045–11050.
10. Nozaki, K., O. Onodera, H. Takano, and S. Tsuji. 2001. Amino acid sequences flanking polyglutamine stretches influence their potential for aggregate formation. *Neuroreport* 12:3357–3364.
11. Rockabrand, E., N. Slepko, A. Pantalone, V. N. Nukala, A. Kazantsev, J. L. Marsh, P. G. Sullivan, J. S. Steffan, S. L. Sensi, and L. M. Thompson. 2007. The first 17 amino acids of huntingtin modulate its sub-cellular localization, aggregation and effects on calcium homeostasis. *Hum. Mol. Genet.* 16:61–77.
12. Chai, Y., L. Wu, J. D. Griffin, and H. L. Paulson. 2001. The role of protein composition in specifying nuclear inclusion formation in polyglutamine disease. *J. Biol. Chem.* 276:44889–44897.
13. Cooper, J. K., G. Schilling, M. F. Peters, W. J. Herring, A. H. Sharp, Z. Kaminsky, J. Masone, F. A. Khan, M. Delaney, D. R. Borchelt, V. L. Dawson, T. M. Dawson, and C. A. Ross. 1998. Truncated N-terminal

- fragments of huntingtin with expanded glutamine repeats form nuclear and cytoplasmic aggregates in cell culture. *Hum. Mol. Genet.* 7:783–790.
14. Haacke, A., S. A. Broadley, R. Boteva, N. Tzvetkov, F. U. Hartl, and P. Breuer. 2006. Proteolytic cleavage of polyglutamine-expanded ataxin-3 is critical for aggregation and sequestration of non-expanded ataxin-3. *Hum. Mol. Genet.* 15:555–568.
  15. Li, M., E. S. Chevalier-Larsen, D. E. Merry, and M. I. Diamond. 2007. Soluble androgen receptor oligomers underlie pathology in a mouse model of spinobulbar muscular atrophy. *J. Biol. Chem.* 282:3157–3164.
  16. Perutz, M. F. 1999. Glutamine repeats and neurodegenerative diseases: molecular aspects. *Trends Biochem. Sci.* 24:58–63.
  17. Chow, M. K., A. M. Ellisdon, L. D. Cabrita, and S. P. Bottomley. 2004. Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. *J. Biol. Chem.* 279:47643–47651.
  18. Chow, M. K., H. L. Paulson, and S. P. Bottomley. 2004. Destabilization of a non-pathological variant of ataxin-3 results in fibrillogenesis via a partially folded intermediate: a model for misfolding in polyglutamine disease. *J. Mol. Biol.* 335:333–341.
  19. Ignatova, Z., and L. M. Gierasch. 2006. Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent  $\beta$  barrel protein. *J. Biol. Chem.* 281:12959–12967.
  20. Tanaka, M., I. Morishima, T. Akagi, T. Hashikawa, and N. Nukina. 2001. Intra- and intermolecular  $\beta$ -pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases. *J. Biol. Chem.* 276:45470–45475.
  21. Chow, M. K., J. P. Mackay, J. C. Whisstock, M. J. Scanlon, and S. P. Bottomley. 2004. Structural and functional analysis of the Josephin domain of the polyglutamine protein ataxin-3. *Biochem. Biophys. Res. Commun.* 322:387–394.
  22. Masino, L., G. Nicastro, R. P. Menon, F. Dal Piaz, L. Calder, and A. Pastore. 2004. Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3. *J. Mol. Biol.* 344:1021–1035.
  23. Bhattacharyya, A., A. K. Thakur, V. M. Chellgren, G. Thiagarajan, A. D. Williams, B. W. Chellgren, T. P. Creamer, and R. Wetzel. 2006. Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* 355:524–535.
  24. Faux, N. G., S. P. Bottomley, A. M. Lesk, J. A. Irving, J. R. Morrison, M. G. de la Banda, and J. C. Whisstock. 2005. Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res.* 15:537–551.
  25. Cabrita, L. D., W. Dai, and S. P. Bottomley. 2006. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* 6:12.
  26. Peters, M. F., and C. A. Ross. 1999. Preparation of human cDNAs encoding expanded polyglutamine repeats. *Neurosci. Lett.* 275:129–132.
  27. Studier, F. W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41:207–234.
  28. Marley, J., M. Lu, and C. Bracken. 2001. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR.* 20:71–75.
  29. Lobley, A., L. Whitmore, and B. A. Wallace. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics.* 18:211–212.
  30. Nagai, Y., T. Tucker, H. Ren, D. J. Kenan, B. S. Henderson, J. D. Keene, W. J. Strittmatter, and J. R. Burke. 2000. Inhibition of polyglutamine protein aggregation and cell death by novel peptides identified by phage display screening. *J. Biol. Chem.* 275:10437–10442.
  31. Ellisdon, A. M., B. Thomas, and S. P. Bottomley. 2006. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J. Biol. Chem.* 281:16888–16896.
  32. Sattler, M., J. Schleucher, and C. Griesinger. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34:93–158.
  33. Kay, L. E., P. Keifer, and P. Saarinen. 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* 114:10663–10665.
  34. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6:277–293.
  35. Herrmann, T., P. Guntert, and K. Wuthrich. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 319:209–227.
  36. Cornilescu, G., F. Delaglio, and A. Bax. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR.* 13:289–302.
  37. Schwieters, C. D., J. J. Kuszewski, N. Tjandra, and G. M. Clore. 2003. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* 160:65–73.
  38. Hwang, T. L., P. C. van Zijl, and S. Mori. 1998. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR.* 11:221–226.
  39. Bottomley, S. P., A. G. Popplewell, M. Scawen, T. Wan, B. J. Sutton, and M. G. Gore. 1994. The stability and unfolding of an IgG binding protein based upon the B domain of protein A from *Staphylococcus aureus* probed by tryptophan substitution and fluorescence spectroscopy. *Protein Eng.* 7:1463–1470.
  40. Sato, S., T. L. Religa, V. Daggett, and A. R. Fersht. 2004. Testing protein-folding simulations by experiment: B domain of protein A. *Proc. Natl. Acad. Sci. USA.* 101:6952–6956.
  41. Torigoe, H., I. Shimada, A. Saito, M. Sato, and Y. Arata. 1990. Sequential <sup>1</sup>H NMR assignments and secondary structure of the B domain of staphylococcal protein A: structural changes between the free B domain in solution and the Fc-bound B domain in crystal. *Biochemistry.* 29:8787–8793.
  42. Klein, F. A., A. Pastore, L. Masino, G. Zeder-Lutz, H. Nierengarten, M. Oulad-Abdelghani, D. Altschuh, J. L. Mandel, and Y. Trotter. 2007. Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. *J. Mol. Biol.* 371:235–244.
  43. Ellisdon, A. M., M. C. Pearce, and S. P. Bottomley. 2007. Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. *J. Mol. Biol.* 368:595–605.
  44. Snell, R. G., J. C. MacMillan, J. P. Cheadle, I. Fenton, L. P. Lazarou, P. Davies, M. E. MacDonald, J. F. Gusella, P. S. Harper, and D. J. Shaw. 1993. Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nat. Genet.* 4:393–397.
  45. Ladurner, A. G., and A. R. Fersht. 1997. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* 273:330–337.
  46. Booth, D. R., M. Sunde, V. Bellotti, C. V. Robinson, W. L. Hutchinson, P. E. Fraser, P. N. Hawkins, C. M. Dobson, S. E. Radford, C. C. Blake, and M. B. Pepys. 1997. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature.* 385:787–793.
  47. Hammarstrom, P., X. Jiang, A. R. Hurshman, E. T. Powers, and J. W. Kelly. 2002. Sequence-dependent denaturation energetics: a major determinant in amyloid disease diversity. *Proc. Natl. Acad. Sci. USA.* 99(Suppl 4):16427–16432.
  48. Horwich, A. 2002. Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J. Clin. Invest.* 110:1221–1232.
  49. Liemann, S., and R. Glockshuber. 1999. Influence of amino acid substitutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein. *Biochemistry.* 38:3258–3267.
  50. Swietnicki, W., R. B. Petersen, P. Gambetti, and W. K. Surewicz. 1998. Familial mutations and the thermodynamic stability of the recombinant human prion protein. *J. Biol. Chem.* 273:31048–31052.



## **Chapter 5**

# **Protein domain architecture determines the aggregation potential of polyglutamine repeats**

Submitted for publication to *Protein Science*.



---

## **Protein domain architecture determines the aggregation potential of polyglutamine repeats**

Amy. L Robertson, Mark A. Bate, Ashley M. Buckle and Stephen P. Bottomley.

Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia 3800.

Correspondence should be addressed to: Stephen P. Bottomley, Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia. Tel: 61-3-99029362; Email: [steve.bottomley@med.monash.edu.au](mailto:steve.bottomley@med.monash.edu.au)

### **Abstract**

The polyglutamine diseases are caused in part by a gain-of-function mechanism of neuronal toxicity involving protein conformational changes that result in the formation and deposition of  $\beta$ -sheet rich aggregates. Recent evidence suggests that the misfolding mechanism is context-dependent, and that properties of the host protein, including the domain architecture and location of the repeat tract, can modulate aggregation. We utilized variants of a model polyQ-containing protein to systematically determine the effect of the location and number of flanking folded domains on polyQ aggregation. Our data show that a polyQ tract present between two domains has a significantly suppressed aggregation rate compared to a variant with a flexible, untethered polyQ tail. We performed a bioinformatic analysis of the location and context of polyQ repeats in all human polyQ-containing proteins, finding that most polyQ repeats in disease and non-disease associated proteins are located in flexible, untethered protein regions. In light of our experimental evidence, these data suggest a physiological importance of the

terminal polyQ location. We further show experimentally that there is a relationship between the number of flanking domains and polyQ aggregation rate. We propose that domain repeats arise through genetic duplication as an evolutionary mechanism to avoid aggregation.

---

**Introduction**

Expansion of a polyglutamine(polyQ)-encoding repetitive CAG sequence is the genetic mechanism leading to neurodegeneration in nine identified diseases, including Huntington's disease (HD) and the spinocerebellar ataxias (SCA) 1, 2, 3, 6, 7, and 17. Although the precise molecular details remain controversial, it is commonly thought that polyQ expansion triggers a toxic gain-of-function disease mechanism; resulting in a conformational change within the protein and the formation and neuronal deposition of  $\beta$ -sheet rich amyloid-like fibrils (Perutz et al. 1994; Chen et al. 2002). Disease results only when polyQ expansion occurs within a specific subset of proteins, and the pathogenic threshold varies within the group of disease proteins, for example a 37 glutamine repeat leads to HD, whilst a repeat of 45 or greater is necessary for SCA3 pathogenesis (Goto et al. 1997; Padiath et al. 2005; Li et al. 2006). Many other neuronally-expressed human proteins not associated with disease contain polyQ repeats (Marcotte et al. 1999; Faux et al. 2005). For example, a Q40 repeat is the normal allele present in the non-disease protein, forkhead box P2 transcription factor (Margolis et al. 1997; Mizutani et al. 2007). Together, this evidence suggests that polyQ behaviour depends upon host protein context and that the gain-of-function mechanism is context-dependent. This raises the question of which protein characteristics modulate disease susceptibility and the ability of polyQ-containing proteins to form fibrils.

Proteins are under evolutionary pressure to avoid causing disease (Rousseau et al. 2006b). Proteins have therefore evolved to oppose aggregation and several host protein characteristics are thought to modulate aggregation (Rousseau et al. 2006a). Amyloidogenic motifs may be buried within the native conformation of a protein, requiring the protein to undergo a local or global conformational change to expose the amyloidogenic region, therefore aggregation-prone protein variants/mutants often have

compromised stability (Hurle et al. 1994; Swietnicki et al. 1998; Liemann and Glockshuber 1999; Horwich 2002). Gatekeeper sequences flanking amyloidogenic regions, rich in charged residues and prolines, can also oppose aggregation (Rousseau et al. 2006a; Rousseau et al. 2006b). The electrostatic charge and presence of prolines in regions flanking polyQ repeats modulate conformation and aggregation propensity (Bhattacharyya et al. 2006; Walters and Murphy 2009). For example, a recent structural description of a Htt exon1 peptide demonstrated that the polyQ tract can adopt the structure of the flanking regions (Kim et al. 2009).

The global protein architecture is also thought to be important for aggregation. Pathogenic polyQ tracts are generally located in flexible protein regions N- or C-terminal to a structured domain (Table 1) (Faux et al. 2005). Proteolytic cleavage is also an initiator of a number of amyloid diseases suggesting that the host protein can protect against aggregation (Stefani and Dobson 2003). It is thought that proteolysis is involved in the pathogenic cascades leading to HD, SCA3, SCA7 and DRPLA (DiFiglia et al. 1997; Wellington et al. 1998; Ellerby et al. 1999). Together these data suggest that domain architecture can modulate polyQ aggregation. Recombinant forms of the disease proteins are difficult to obtain, therefore limiting the *in vitro* comparison of aggregation for full-length disease proteins and cleaved fragments.

In this study we assessed the effect of flanking folded domains on polyQ aggregation using a model polyQ-containing protein. We asked whether the location and number of domains surrounding the polyQ tract modulate aggregation. We investigated the significance of repeat tract location in the context of a polyQ-containing model protein, and our experimental data suggest that polyQ location and domain architecture affect the rate of polyQ-dependent fibril formation. As polyQ-encoding repeats are frequent within the human genome we addressed whether repeat

location and protein domain architecture affect the disease susceptibility of human polyQ proteins.

---

**Results****Interdomain polyQ location retards aggregation**

In this study we tested the hypothesis that the presence of flanking domains affects the rate of polyQ aggregation. We utilized variants of a polyQ-containing model protein, comprising the small, three-helical *Staphylococcus aureus* Protein A B domain (SpA) (Bottomley et al. 1994; Sato et al. 2004). We have previously shown that a variant of this protein with an expanded polyQ fused to the C-terminus of SpA forms ThT-reactive fibrils at a repeat-length dependent rate by a mechanism involving conformational changes only within the polyQ tract (Robertson et al. 2008). We firstly investigated the effect of repeat tract location on polyQ-dependent aggregation. We used SpA-polyQ variants with the polyQ present between two SpA domains, which we have termed interdomain (id) variants (Figure 1A). At a concentration of 65  $\mu$ M SpA-cQ48 rapidly formed ThT-reactive aggregates, with no detectable lag phase, and a reaction half-life of approximately 3h (Figure 1B). When the Q48 tract was present between two SpA domains in the SpA-idQ48 construct, the rate of aggregation was considerably slowed, with a half-life of 130h (Figure 1B). At protein concentrations lower than this, aggregation occurred too slowly for adequate analyses. We expanded the length of the polyQ tract in the interdomain context to make an SpA-idQ64 construct. This variant formed aggregates at a faster rate than SpA-idQ48, with a half-life of 12h (Figure 1B). This was still significantly slower than the rate of SpA-cQ48 aggregation. Considering there is a repeat-length dependent association with aggregation rate, these data indicate that the sequence context of the polyQ tract significantly modulates the rate of polyQ aggregation.

---

**Domain duplication slows polyQ aggregation**

Many poly amino acid-containing proteins are transcription factors that contain multiple repeats of a single domain (Faux et al. 2005). To determine the relationship between the number of duplicated domains and aggregation rate we engineered protein variants comprising one, two or three SpA domains fused to the N-terminus of a Q45 repeat, denoted SpA-cQ45, SpA<sub>2</sub>-cQ45 and SpA<sub>3</sub>-cQ45 respectively (Figure 2A). The half-life of ThT-reactive aggregation increased as the number of N-terminally fused SpA domains increased, with half-lives of 9h, 61h and 87h for the variants with one, two and three SpA domains respectively (Figure 2B). The prolonged lag phase and aggregation half-lives indicate a slowed rate of nucleation, however it is also clear that the presence of multiple SpA domains inhibits the elongation phase of the nucleation-dependent kinetics, as indicated by the decreased gradient of the aggregate growth phase with increasing number of SpA domains (Figure 2B).

**Bioinformatic analysis of human polyQ-containing proteins**

Our experimental data show that the rate of polyQ aggregation is linked to the location and number of flanking domains. To investigate whether polyQ location and protein domain architecture are factors that determine the involvement of polyQ-containing proteins in disease, we searched NCBI's Non-Redundant protein database (NCBI NR) for all human proteins containing a polyQ repeat  $\geq 7$ . After sequence analysis we derived a set of 128 non-redundant human polyQ-containing proteins.

Using Pfam (Finn et al. 2008) we assessed the domain architectures of the 128 proteins to determine the location of the polyQ tract relative to known domains. We determined whether the polyQ repeats were N- or C-terminal to a domain, between domains (interdomain), or within a domain (middomain).

There is a high frequency of N-terminal polyQ repeats across all polyQ proteins (Table 2). We could not experimentally determine the aggregation rate of variants of our model protein with N-terminal polyQ repeats due to difficulties in protein production, however the polyQ tract in the C-terminal polyQ proteins has a similar local environment.

There is a higher proportion of interdomain polyQ in the non-disease group compared with the disease-associated proteins. There is a significantly higher frequency of middomain polyQ in disease compared to non-disease proteins (Table 2). Androgen receptor and atrophin-1 both have their polyQ present within a defined Pfam domain. Only 3% of non-disease proteins have a middomain polyQ tract.

As the experimental data showed that multiple SpA domains fused to the N-terminus of a polyQ tract can modulate aggregation, we investigated all human polyQ proteins for the presence of multiple/tandem repeats of domains. There is no significant difference in the frequency of domain repeats between the disease and non-disease polyQ proteins (Table 2). The disease proteins, ataxin-3, TATA-binding protein and CACNA1A all contain at least two repeats of a single domain.

Interestingly, 47% of proteins that contain multiple/tandem repeats of domains also have an interdomain polyQ location. In light of our model protein data showing significant retardation of aggregation when the polyQ is flanked at the N- and C-termini by SpA domains, we propose that domain repeats arise through genetic duplication as an evolutionary mechanism to avoid aggregation.



---

**Discussion**

The polyQ diseases are regarded as gain-of-function diseases, suggesting that an expanded polyQ tract is the only requirement for pathogenesis. In several cases addition of ectopic pathological length polyQ repeats to a protein does induce aggregation and/or cytotoxicity (Ordway et al. 1997; Robertson et al. 2008). Evidence that the polyQ gain-of-function is context-dependent, however, suggests that properties extrinsic to the polyQ can modulate aggregation. Because of difficulties in characterizing the disease proteins, we utilized a polyQ model protein to determine the effects of domain architecture on polyQ aggregation. The results of our study directly show that the number and location of domains relative to the polyQ tract have a significant effect on the rate of aggregation.

We showed that an interdomain polyQ variant has a significantly suppressed rate of aggregation compared to a protein without a second folded SpA domain C-terminal to the polyQ. In a previous study we showed that the structure, stability and dynamics of the SpA domain are unaffected when fused to a polyQ tract, suggesting that the behaviour of the polyQ tract is critical to aggregation (Robertson et al. 2008). In light of this, the slowed aggregation kinetics of the interdomain polyQ variant may be due to perturbations to the conformational freedom and intermolecular associations of the polyQ region.

The kinetic mechanism of polyQ aggregation is thought to be initiated by misfolding within the monomer (Chen et al. 2002). The prolonged lag phase for aggregation of the interdomain polyQ suggests that the rate of nucleation is slower than SpA-cQ48. The additional SpA domain may make the misfolded nucleus less kinetically, thermodynamically or sterically accessible. The presence of folded domains suppresses aggregation nucleation of another amyloid-forming protein, Sup35p

(Krammer et al. 2008). The rate of elongation is also much slower for SpA-idQ48 compared to SpA-cQ48 suggesting that steric hindrance may also affect the propensity for intermolecular interactions leading to aggregation.

Alternatively, the increased protein size caused by additional SpA domains could be an inhibitory factor. The slowed aggregation of SpA<sub>2</sub>-cQ45 and SpA<sub>3</sub>-cQ45 compared with the single domain variant suggests that the increased molecular weight and/or total electrostatic charge/solubility caused by additional domains is an inhibitory factor. However, this does not completely account for the slower aggregation of SpA-idQ48 compared to SpA-cQ48, suggesting that the interdomain location of the polyQ is largely responsible for modulating aggregation.

Of the nine disease proteins, only ataxin-3 has an interdomain polyQ tract. However, the ataxin-3 isoform commonly observed in disease plaques is a truncated variant lacking the C-terminal ubiquitin interacting motif (UIM), and proteolytic cleavage is also thought to be involved in SCA3 pathogenesis (Schmidt et al. 1998; Wellington et al. 1998). In the full-length context the C-terminal UIM may suppress aggregation. Our database analysis showed that 21 of the 119 non-disease polyQ proteins have an interdomain polyQ (Table 2). The interdomain location of the polyQ may be a factor that suppresses the disease potential of these proteins.

Our experimental data shows a strong correlation between the number of duplicated domains and aggregation rate (Figure 2B). The database analysis indicates that the frequency of domain repeats is similar for disease and non-disease polyQ-containing proteins. Although the frequency of interdomain polyQ repeats is low, a high proportion of the polyQ in these proteins are present between repeated domains. These data suggest that genetic duplication of domains leading to an interdomain polyQ location may be a conserved mechanism to avoid aggregation.

Our database analysis shows that there are at least 128 human proteins that contain a polyQ repeat  $\geq 7$ . There are a growing number of polyQ diseases, and it is possible that there are some unidentified disease-causing proteins. Our data could therefore be used as a tool to identify other disease proteins. Many of these proteins are transcription factors that are localized to the nucleus and are neuronally expressed, making many of the proteins disease candidates. A previous study has shown that proteins expressed in the brain with a repeat of  $>6\text{CAG}$  have a high likelihood to undergo expansion mutation (Margolis et al. 1997). The analysis shows that most polyQ, irrespective of their involvement in disease, are present in a flexible region of the protein, and not flanked by domains at both termini. In light of our experimental data, this conservation of polyQ location is detrimental in terms of aggregation and disease-causing potential. Together these data indicate that other protein and cellular factors are involved in modulation of aggregation and disease. These data suggest that evolutionary pressure on polyQ-containing proteins is more a consequence of the functional requirements of the polyQ.

## **Conclusions**

The first polyQ disease was identified in 1991, therefore, the polyQ field is relatively young and it is likely that there are unidentified disease-causing proteins. Early on it was established that an expanded polyQ repeat leads to a gain-of-function misfolding and aggregation mechanism. Only recently has the protein context been considered to have a role in aggregation. A number of studies have focused on the modulatory role of the local sequence on polyQ aggregation. Prior to this investigation the modulatory effect of flanking folded domains had not been experimentally investigated. Our model protein data have shown that the rate of polyQ aggregation correlates with the number

---

and location of flanking domains. Our bioinformatic analysis of the human proteome shows that most polyQ repeats are located in flexible, untethered protein regions N- or C-terminal to a structured domain. As our experimental data suggest that this location enhances aggregation propensity, these data combined suggest a functional importance of repeat location. The data presented here show that polyQ location is one element that can modulate a protein's aggregation propensity, however, any protein containing a polyQ tract may be prone to aggregation/disease if the protein is susceptible to proteolytic cleavage.

## **Materials and Methods**

### **Plasmid construction**

SpA-idQ48 and SpA-idQ64 were constructed from the templates SpA-cQ48 pET21a and SpA-cQ64 pET21a (manuscript in preparation). The stop codon was removed from SpA-cQ48 and SpA-cQ64 by Quikchange™ site-directed mutagenesis using the complementary primers; 5'-GCGGCCCCGGGGCAGGATCCGAATTC-3' and 5'-GAATTCGGATCCTGCCCCGGGCGC-3'. The SpA domain was PCR amplified using the forward primer 5'-CAGGGCCACGTGGGATCCGATAACAAAT-3', which introduced a *Bam*H1 site; and reverse primer 5'-CTCGAATTCCTCGAGTCACCCGGGCGCGGTT-3', introducing a *Xho*1 site. SpA was then subcloned into the templates by ligation dependent cloning using *Bam*H1 and *Xho*1 restriction sites. SpA<sub>2</sub>-cQ45 was constructed using the SpA-cQ45 pET21a template. A *Kpn*1 restriction site was inserted into SpA-cQ45 by site-directed mutagenesis using the complementary primers 5'-GCCACGTGGACGTCGGTACCGATAACAAATTTAATAA-3' and 5'-TTATTAAATTTGTTATCGGTACCGACGTCCACGTGGC-3'. SpA was PCR amplified using a forward primer containing a *Pml*1 site, 5'-GGGCGGCAGGGCCACGTGGACGTCGATAAC-3', and reverse primer containing a *Kpn*1 site, 5'-GGGCGGCAGGGCCACGTGGACGTCGATAAC-3' followed by restriction cloning into SpA-cQ45 pET21a. SpA<sub>3</sub>-cQ45 was constructed using the SpA<sub>2</sub>-cQ45 pET21a template. An *Spe*1 site was inserted into SpA<sub>2</sub>-cQ45 pET21a by site-directed mutagenesis using the primers, 5'-CACGTGGACGTCACTAGTGATAACAAATTTA-3' and 5'-TAAATTTGTTATCACTAGTGACGTCCACGTG-3'. SpA was amplified using the *Pml*1-containing primer as above and a reverse primer incorporating an *Spe*1 site, 5'-

CCCGGGCCGCGGACTAGTTTTTCGGCGCCTG-3', restriction digested with *Pml*I and *Spe*I and ligated with SpA<sub>2</sub>-cQ45 pET21a.

### **Protein preparation**

Proteins were prepared according to the previously described method for SpA-polyQ proteins (Robertson et al. 2008).

### **Aggregation experiments**

Aggregation reactions were performed at concentrations of 20  $\mu$ M, 50  $\mu$ M and 65  $\mu$ M. Reactions were carried out in 384-well clear-bottom plates and aggregation was monitored by ThT fluorescence, by measuring the emission intensity at 490 nm ( $\lambda_{\text{ex}}$  = 450 nm) on a BMG Laboratories FLUOstar Optima fluorescence plate reader. All reactions were performed in 20 mM sodium acetate (pH 5.0), 100 mM NaCl, 2mM PMSF, 20  $\mu$ M ThT and 10 % glycerol.

### **Bioinformatic analysis of human polyQ-containing proteins**

We searched NCBI's Non-Redundant protein database (NCBI NR) for all human proteins containing a polyQ repeat  $\geq 7$ . The NCBI NR contains many versions of the same protein, which created bias in the statistical analysis of polyQ location data. We simplified the analysis by indentifying protein variants/isoforms and using only the longest protein isoforms (which we termed "master sequences"), therefore eliminating splice variants/protein fragments. Multiple variants/isoforms of each protein were firstly crudely identified by comparing the protein sequence following the polyQ chains, the original sequences were then ran through the BLASTClust (Altschul et al. 1990), FORCE (Wittkop et al. 2007), MCL (Dongen 2000), and HomoClust algorithms (Chen

---

et al. 2006), and the variants/isoforms were adjusted as necessary. The crude identification used the 10 amino acids immediately after the polyQ chain as a “search string”, any sequence that had the 10 amino acids immediately following it’s own polyQ chain was presumed to have homology with that sequence. The homology groups were confirmed by analyzing the data using the above algorithms. This yielded a total of 128 master sequences, from an original dataset of >700 polyQ-containing human protein sequences.

**References**


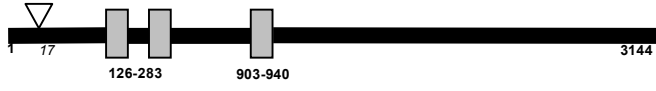
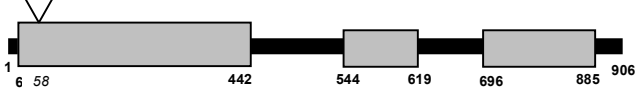

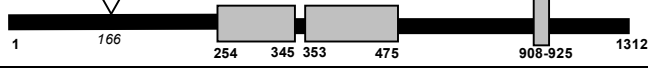

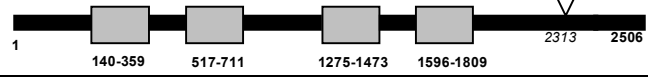
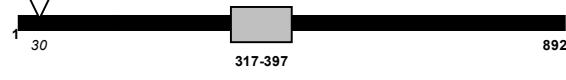

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J Mol Biol* **215**(3): 403-410.
- Bhattacharyya, A., Thakur, A.K., Chellgren, V.M., Thiagarajan, G., Williams, A.D., Chellgren, B.W., Creamer, T.P., and Wetzel, R. 2006. Oligoproline effects on polyglutamine conformation and aggregation. *J Mol Biol* **355**(3): 524-535.
- Bottomley, S.P., Popplewell, A.G., Scawen, M., Wan, T., Sutton, B.J., and Gore, M.G. 1994. The stability and unfolding of an IgG binding protein based upon the B domain of protein A from *Staphylococcus aureus* probed by tryptophan substitution and fluorescence spectroscopy. *Protein Eng* **7**(12): 1463-1470.
- Chen, C.-Y., Chung, W.-C., and Su, C.-T. 2006. Exploiting homogeneity in protein sequence clusters for construction of protein family hierarchies. *Pattern Recogn* **39**(12): 2356-2369.
- 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**(18): 11884-11889.
- DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., and Aronin, N. 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**(5334): 1990-1993.
- Dongen, S.v. 2000. Graph Clustering by Flow Simulation. Vol PhD. Univeristy of Utrecht.
- Ellerby, L.M., Hackam, A.S., Propp, S.S., Ellerby, H.M., Rabizadeh, S., Cashman, N.R., Trifiro, M.A., Pinsky, L., Wellington, C.L., Salvesen, G.S. et al. 1999. Kennedy's disease: caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. *J Neurochem* **72**(1): 185-195.
- Faux, N.G., Bottomley, S.P., Lesk, A.M., Irving, J.A., Morrison, J.R., de la Banda, M.G., and Whisstock, J.C. 2005. Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res* **15**(4): 537-551.
- Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L. et al. 2008. The Pfam protein families database. *Nucleic Acids Res* **36**(Database issue): D281-288.
- Goto, J., Watanabe, M., Ichikawa, Y., Yee, S.B., Ihara, N., Endo, K., Igarashi, S., Takiyama, Y., Gaspar, C., Maciel, P. et al. 1997. Machado-Joseph disease gene products carrying different carboxyl termini. *Neurosci Res* **28**(4): 373-377.
- Horwich, A. 2002. Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J Clin Invest* **110**(9): 1221-1232.
- Hurle, M.R., Helms, L.R., Li, L., Chan, W., and Wetzel, R. 1994. A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proc Natl Acad Sci U S A* **91**(12): 5446-5450.
- Kim, M.W., Chelliah, Y., Kim, S.W., Otwinowski, Z., and Bezprozvanny, I. 2009. Secondary structure of Huntingtin amino-terminal region. *Structure* **17**(9): 1205-1212.
- Krammer, C., Kremmer, E., Schatzl, H.M., and Vorberg, I. 2008. Dynamic interactions of Sup35p and PrP prion protein domains modulate aggregate nucleation and seeding. *Prion* **2**(3): 99-106.
- Li, W., Serpell, L.C., Carter, W.J., Rubinsztein, D.C., and Huntington, J.A. 2006. Expression and characterization of full-length human huntingtin, an elongated HEAT repeat protein. *J Biol Chem* **281**(23): 15916-15922.



- Liemann, S. and Glockshuber, R. 1999. Influence of amino acid substitutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein. *Biochemistry* **38**(11): 3258-3267.
- Marcotte, E.M., Pellegrini, M., Yeates, T.O., and Eisenberg, D. 1999. A census of protein repeats. *J Mol Biol* **293**(1): 151-160.
- Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S., Breschel, T.S., Stine, O.C., Callahan, C., McInnis, M.G., and Ross, C.A. 1997. cDNAs with long CAG trinucleotide repeats from human brain. *Hum Genet* **100**(1): 114-122.
- Mizutani, A., Matsuzaki, A., Momoi, M.Y., Fujita, E., Tanabe, Y., and Momoi, T. 2007. Intracellular distribution of a speech/language disorder associated FOXP2 mutant. *Biochem Biophys Res Commun* **353**(4): 869-874.
- Ordway, J.M., Tallaksen-Greene, S., Gutekunst, C.A., Bernstein, E.M., Cearley, J.A., Wiener, H.W., Dure, L.S.t., Lindsey, R., Hersch, S.M., Jope, R.S. et al. 1997. Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell* **91**(6): 753-763.
- Padiath, Q.S., Srivastava, A.K., Roy, S., Jain, S., and Brahmachari, S.K. 2005. Identification of a novel 45 repeat unstable allele associated with a disease phenotype at the MJD1/SCA3 locus. *Am J Med Genet B Neuropsychiatr Genet* **133B**(1): 124-126.
- 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**(12): 5355-5358.
- Robertson, A.L., Horne, J., Ellisdon, A.M., Thomas, B., Scanlon, M.J., and Bottomley, S.P. 2008. The structural impact of a polyglutamine tract is location-dependent. *Biophys J* **95**(12): 5922-5930.
- Rousseau, F., Schymkowitz, J., and Serrano, L. 2006a. Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* **16**(1): 118-126.
- Rousseau, F., Serrano, L., and Schymkowitz, J.W. 2006b. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* **355**(5): 1037-1047.
- Sato, S., Religa, T.L., Daggett, V., and Fersht, A.R. 2004. Testing protein-folding simulations by experiment: B domain of protein A. *Proc Natl Acad Sci U S A* **101**(18): 6952-6956.
- Schmidt, T., Landwehrmeyer, G.B., Schmitt, I., Trottier, Y., Auburger, G., Laccone, F., Klockgether, T., Volpel, M., Epplen, J.T., Schols, L. et al. 1998. An isoform of ataxin-3 accumulates in the nucleus of neuronal cells in affected brain regions of SCA3 patients. *Brain Pathol* **8**(4): 669-679.
- Stefani, M. and Dobson, C.M. 2003. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* **81**(11): 678-699.
- Swietnicki, W., Petersen, R.B., Gambetti, P., and Surewicz, W.K. 1998. Familial mutations and the thermodynamic stability of the recombinant human prion protein. *J Biol Chem* **273**(47): 31048-31052.
- Walters, R.H. and Murphy, R.M. 2009. Examining Polyglutamine Peptide Length: A Connection between Collapsed Conformations and Increased Aggregation. *J Mol Biol*.
- Wellington, C.L., Ellerby, L.M., Hackam, A.S., Margolis, R.L., Trifiro, M.A., Singaraja, R., McCutcheon, K., Salvesen, G.S., Propp, S.S., Bromm, M. et al. 1998. Caspase cleavage of gene products associated with triplet expansion

- 
- disorders generates truncated fragments containing the polyglutamine tract. *J Biol Chem* **273**(15): 9158-9167.
- Wittkop, T., Baumbach, J., Lobo, F.P., and Rahmann, S. 2007. Large scale clustering of protein sequences with FORCE -A layout based heuristic for weighted cluster editing. *BMC Bioinformatics* **8**: 396.

**Table 1.** The domain architecture of the nine polyQ disease proteins.

| Disease       | Protein           | Protein architecture*  |
|---------------|-------------------|--|
| DRPL<br>A     | Atrophin-1        |    |
| HD            | Huntingtin        |    |
| SBMA          | Androgen Receptor |    |
| SCA1          | Ataxin-1          |    |
| SCA2          | Ataxin-2          |    |
| SCA3<br>(MJD) | Ataxin-3          |    |
| SCA6          | CACNA1A           |   |
| SCA7          | Ataxin-7          |  |
| SCA17         | TBP               |   |

Abbreviations: DRPLA- dentatorubropallidoluysian atrophy; HD- Huntington's disease; SBMA- Spinobulbar muscular atrophy; SCA- spinocerebellar ataxia; MJD- Machado Joseph disease; CACNA1A- alpha1 voltage-dependent calcium channel; TBP- TATA-box binding protein.

\*Protein architectures and domain boundaries were identified using Pfam. Triangles indicate the position of the polyQ (italicized numbers indicate the position of the first glutamine in the repeat tract). Figures are not to scale. Black regions indicate structurally undefined regions (the domain architecture of atrophin-1 has not been described) and grey boxes indicate defined domains.

**Table 2:** Location of the polyQ tract in all human polyQ-containing proteins.

|                          | <b>All PolyQ: 128 proteins</b> | <b>Non-Disease PolyQ: 119 proteins</b> | <b>Disease PolyQ: 9 proteins</b> |
|--------------------------|--------------------------------|--|----------------------------------|
| <b>N-Terminal PolyQ</b>  | 37 (29%)                       | 33 (28%)                               | 4 (44%)                          |
| <b>C-Terminal PolyQ</b>  | 36 (28%)                       | 35 (29%)                               | 1 (11%)                          |
| <b>Interdomain PolyQ</b> | 22 (17%)                       | 21 (18%)                               | 1 (11%)                          |
| <b>Middomain PolyQ</b>   | 5 (4%)                         | 3 (3%)                                 | 2 (22%)                          |
| <b>No-Domain PolyQ</b>   | 27 (21%)                       | 26 (22%)                               | 1 (11%)                          |

N-terminal PolyQ- Proteins with a polyQ repeat N-terminal to a Pfam-defined domain. C-terminal PolyQ- Proteins with a polyQ repeat C-terminal to a Pfam-defined domain. Interdomain polyQ- Proteins with a polyQ repeat between two domains. Middomain polyQ- Proteins with a polyQ repeat within a domain. No-domain polyQ- Proteins containing a polyQ repeat that have no Pfam defined domains.

---

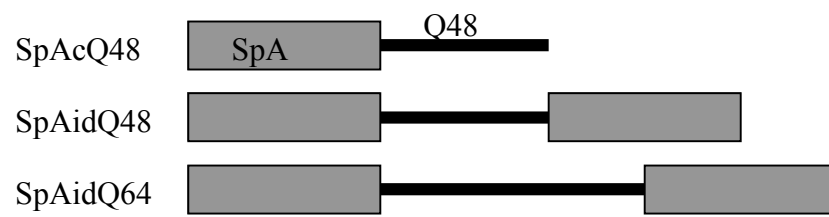
**Figure Legends**

**Figure 1:** A) Schematic diagram of SpA-cQ48, SpA-idQ48 and SpA-idQ64 constructs. Grey boxes represent SpA domains and the black bars represent the polyQ region. B) Aggregation of SpA-cQ48 (black), SpA-idQ48 (grey) and SpA-idQ64 (black, dashed) monitored by ThT fluorescence.

**Figure 2:** A) Schematic diagram of SpA-cQ45, SpA<sub>2</sub>-cQ45 and SpA<sub>3</sub>-cQ45 constructs. Grey boxes represent SpA domains and black bars represent the polyQ region. B) Aggregation of SpA-cQ45 (black), SpA<sub>2</sub>-cQ45 (black, dashed) and SpA<sub>3</sub>-cQ45 (grey) monitored by ThT fluorescence.

Figure 1.

A.



B.

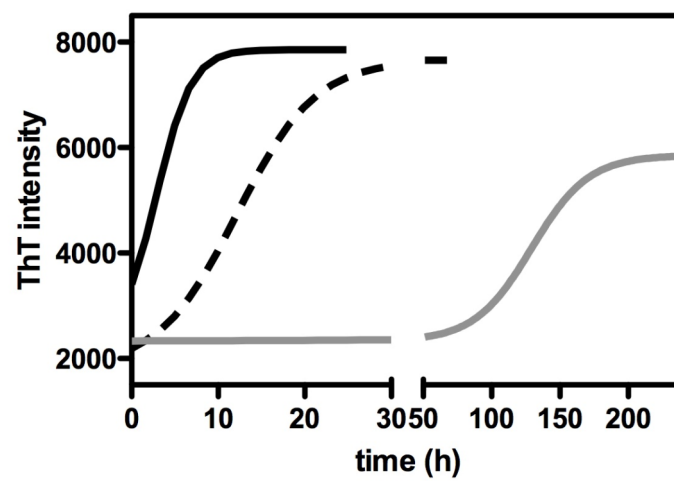
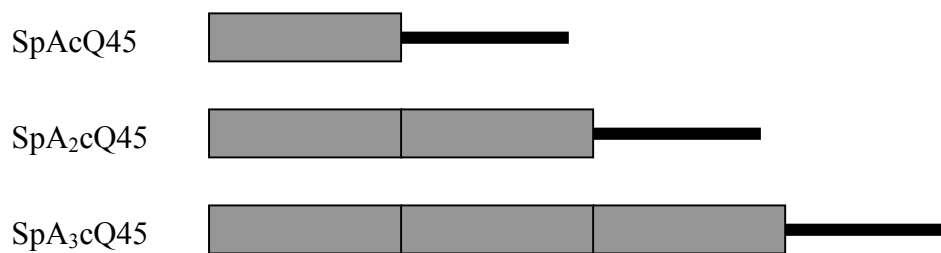
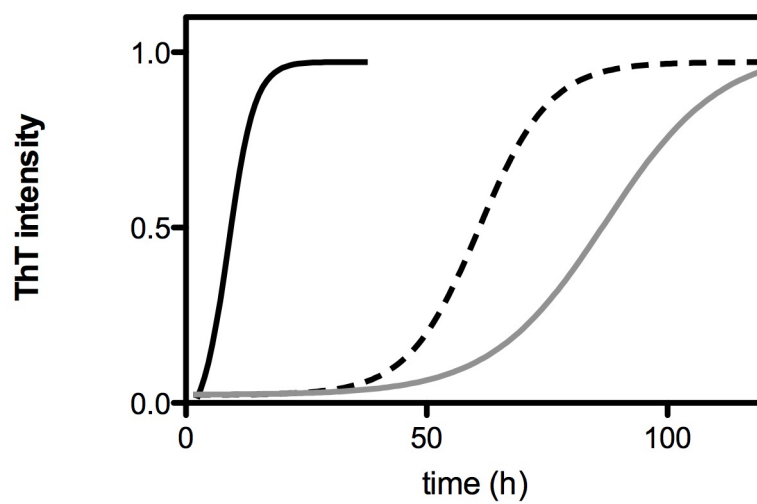


Figure 2.

A.



B.



## **Chapter 6**

### **Small heat-shock proteins inhibit polyglutamine aggregation by interactions with a flanking domain**

Submitted to *Proceedings of the National Academy of Sciences*.



## **Small heat-shock proteins inhibit polyglutamine aggregation by interactions with a flanking domain**

Amy L. Robertson<sup>1</sup>, Stephen J. Headey<sup>2</sup>, Helen M. Saunders<sup>1</sup>, Heath Ecroyd<sup>3\*</sup>, Martin J. Scanlon<sup>2</sup>, John A. Carver<sup>3</sup> and Stephen P. Bottomley<sup>1</sup>.

<sup>1</sup>Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, 3800, Australia, <sup>2</sup> Medicinal Chemistry and Drug Action, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, 3052, Australia and <sup>3</sup>School of Chemistry & Physics, The University of Adelaide, Adelaide, South Australia, 5005, Australia.

\* Present address – School of Biological Sciences, University of Wollongong, Wollongong, New South Wales, 2522, Australia.

Correspondence should be addressed to: Stephen P. Bottomley, Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia. Tel: 61-3-99029362; Email: [steve.bottomley@med.monash.edu.au](mailto:steve.bottomley@med.monash.edu.au)

**Abstract**

Small heat-shock proteins (sHsps) are molecular chaperones which play an important protective role against cellular protein misfolding by interacting with partially-unfolded proteins on their off-folding pathway, thereby preventing their aggregation. Polyglutamine (polyQ) repeat expansion leads to the formation of fibrillar protein aggregates and neuronal cell death in nine diseases, including Huntington's disease (HD) and the spinocerebellar ataxias (SCAs) 1,2,3,6,7, and 17. There is evidence that sHsps have a role in suppression of polyQ-induced neurodegeneration; for example, the sHsp alphaB-crystallin ( $\alpha$ B-c) has been identified as a suppressor of SCA3 toxicity in a *Drosophila* model. However, the molecular mechanism for this suppression is unknown. In this study we tested the ability of  $\alpha$ B-c to suppress the aggregation of a polyQ protein. We found that  $\alpha$ B-c does not inhibit the formation of SDS-insoluble polyQ fibrils. We further tested the effect of  $\alpha$ B-c on the aggregation of ataxin-3 (at3), a polyQ protein that aggregates via a two-stage aggregation mechanism. The first stage involves association of the N-terminal Josephin domain followed by a second stage of polyQ-mediated interactions and the formation of SDS-resistant mature fibrils. Our data show that  $\alpha$ B-c potently inhibits the first stage of at3 aggregation, however, the second polyQ-dependent stage can still proceed. Moreover, the rate of at3 fibril formation is significantly reduced by increasing  $\alpha$ B-c concentrations, confirming that initial Josephin aggregation enhances the rate of polyQ intermolecular interactions. Using NMR spectroscopy, we have determined that  $\alpha$ B-c interacts with an extensive region on the surface of the Josephin domain. These data provide an example of a domain/region flanking an amyloidogenic sequence that has a critical role in modulating aggregation of

a polypeptide and plays a role in the interaction with molecular chaperones to prevent this aggregation.

### **Introduction**

Small heat-shock proteins (sHsps) are important in the maintenance of cellular homeostasis. sHsps are induced under stress conditions and interact with partially-unfolded proteins on their off-folding pathways thereby providing a protective mechanism against protein misfolding and aggregation (1). They are present in many species and 10 have been identified in the human proteome [see reviews (1, 2)]. Alpha-crystallin ( $\alpha$ -crystallin) is the most well characterized sHsp. The two isoforms of  $\alpha$ -crystallin,  $\alpha$ A-crystallin ( $\alpha$ A-c) and  $\alpha$ B-crystallin ( $\alpha$ B-c) are a large component of the human eye lens, and  $\alpha$ B-c is expressed in many other cell types including neurons (3, 4). Monomeric  $\alpha$ B-c is approximately 20 kDa in mass and under native conditions it forms a heterogeneous array of multimeric complexes ranging from 160 to 1000 kDa in mass (2). The activity of  $\alpha$ B-c involves hydrophobic interactions between  $\alpha$ B-c and the partially folded protein. Under stress conditions, such as elevated temperature, the activity of  $\alpha$ B-c is enhanced due to a conformational change within the multimeric complex (5-7). Cellular stress is induced in several neurodegenerative diseases, such as Alzheimer disease (8), prion diseases (9), Parkinson disease (4) and Huntington disease (HD) (4), and there is evidence that the activity of  $\alpha$ B-c is increased (4). In addition,  $\alpha$ B-c reduces neurotoxicity in a number of disease models (10, 11). *In vitro* the interaction between  $\alpha$ B-c and a number of aggregating proteins has been characterized. Most substrates aggregate through a nucleation-dependent kinetic mechanism, however the mode of interaction with  $\alpha$ B-c appears to be dependent upon the specific target protein and the type of aggregation, ie. amorphous or fibrillar (12-15). Complex stability is influenced by properties of the target protein, for example the fibril-forming

proteins, apolipoprotein-CII (ApoCII) (12) and amyloid- $\beta$  (A $\beta$ ) (16) transiently interact with  $\alpha$ B-c whilst  $\alpha$ -synuclein (6) and  $\kappa$ -casein (17) form stable non-covalent complexes. Determining the mechanism by which  $\alpha$ B-c interacts with specific aggregation-prone substrates is therefore important and may provide insight into the misfolding pathways leading to toxicity.

Polyglutamine (polyQ) expansion leads to the formation of fibrillar protein aggregates and neuronal cell death in nine diseases, including HD and the spinocerebellar ataxias (SCAs) 1,2,3,6,7, and 17. PolyQ peptides and proteins form fibrillar aggregates by a nucleation-dependent mechanism that is initiated by a monomeric nucleus (18-20). Recent evidence suggests that polyQ protein misfolding not only involves the polyQ tract but other aggregation prone regions within the polyQ proteins (21-23). There is experimental evidence that the proteins, ataxin-3 (at3) (in SCA3), ataxin-1 (in SCA1) and huntingtin (in HD) form fibrillar aggregates by a multi-domain misfolding mechanism in which non-polyQ regions of the protein self-associate before the polyQ tract (21, 22, 24). Our laboratory has previously shown that at3 has a minimal two-stage aggregation mechanism (21). The first stage involves aggregation of the globular N-terminal Josephin domain which precedes the self-association of expanded polyQ segments (21).

sHsp over-expression decreases neuronal toxicity in HD by non-chaperone mechanisms, including suppression of reactive oxygen species (ROS) (25) and stimulation of autophagy (26). In disease models over-expression of  $\alpha$ B-c modulates toxicity in a context-dependent manner.  $\alpha$ B-c did not alter the HD phenotype in a *Drosophila* model (27), however, suppressed SCA3 toxicity in a *Drosophila* model (10). Enhanced suppression was observed when  $\alpha$ B-c was co-expressed with full-length ataxin-3 in comparison with a C-terminal fragment not containing the Josephin

domain (10). These data suggest that non-polyQ protein regions influence sHsp chaperone activity.

In this study we tested the chaperone ability of  $\alpha$ B-c for a polyQ protein (SpA-cQ52) that forms fibrillar aggregates by a mechanism involving only the polyQ region. We found that  $\alpha$ B-c does not inhibit fibrillogenesis in this system. We then tested the effect of  $\alpha$ B-c on the aggregation of at3 harboring a pathological length polyQ tract (at3(Q64)), and a truncated variant comprising only the Josephin domain. Our data show that  $\alpha$ B-c potently inhibits the initial Josephin-dependent stage of at3 aggregation. Using NMR spectroscopy, we determined that  $\alpha$ B-c interacts with specific regions of Josephin. These data provide for the first time, an example and characterization of a domain/region flanking an amyloidogenic sequence that has a critical role in modulation of aggregation by interacting with molecular chaperones.

---

**Results** **$\alpha$ B-crystallin does not inhibit polyQ mediated aggregation.**

We first tested whether  $\alpha$ B-c can prevent polyQ tract mediated aggregation. We co-incubated  $\alpha$ B-c with a polyQ protein (SpA-cQ52), which has previously been shown to form SDS-insoluble aggregates by a mechanism involving only the polyQ tract (28). The rate at which SpA-cQ52 forms ThT-reactive aggregates does not change when incubated with increasing concentrations of  $\alpha$ B-c (Figure 1A). It has been previously shown that molecular chaperones can redirect the aggregation pathway of a protein from amyloid to amorphous aggregation (29, 30). Amorphous aggregates can be differentiated from amyloid by their sensitivity to SDS and appearance when visualized by TEM. So we tested whether SpA-cQ52 incubated with  $\alpha$ B-c forms SDS-insoluble aggregates using a membrane filter trap assay (31, 32). These data demonstrate that the aggregates formed in the presence or absence of  $\alpha$ B-c are insoluble in SDS (Figure 1B). In addition, EM analysis of the aggregates indicated that  $\alpha$ B-c does not affect their fibrillar morphology. Fibrils ranging from 15-18 nm in width and 300-400 nm in length were observed both in the presence and absence of  $\alpha$ B-c (Figure 1C). Together, these data demonstrate that  $\alpha$ B-c does not inhibit or modify the fibrillar aggregation of SpA-cQ52.

 **$\alpha$ B-crystallin inhibits the Josephin-dependent stage of the two-stage ataxin-3 aggregation pathway.**

We have previously reported that full-length ataxin-3 forms aggregates by a two-stage mechanism (20). The aggregates formed in both stages are ThT-reactive and in the absence of glycerol the two stages are characterised by a biphasic time-course (Figure 2A). The first stage is dominated by interactions of the N-terminal Josephin domain, which is an independently aggregation-prone domain (33). Aggregates formed during

the first stage appear as short curvilinear structures and can be solubilized by SDS (21). These aggregates are stabilized in the second stage by interactions of the polyQ regions. Stage two aggregates can be distinguished from stage one aggregates because they are SDS-insoluble and morphologically appear as larger fibrillar structures. We determined the effect of  $\alpha$ B-c on the two-stage aggregation pathway of a pathological length variant of ataxin-3 (at3(Q64)).  $\alpha$ B-c significantly slows the formation of ThT-reactive at3(Q64) aggregates in a concentration-dependent manner (Fig. 2A, Table 1).

The Josephin domain forms short aggregates with a ribbon-like morphology (33), however the aggregation kinetics have not been described. Isolated Josephin forms ThT-reactive aggregates at a much slower rate than at3(Q64) (Figure 2B, Table 1). The half-life of the first ThT-reactive phase of at3(Q64) aggregation represents the formation of Josephin-dependent fibrils (21). The half-life shifts from approximately 9 h in the full-length protein to around 60 h for the isolated Josephin domain (Table 1). These data indicate that the C-terminal polyQ tail in the context of the full-length protein accelerates Josephin domain aggregation.

$\alpha$ B-c has potent inhibitory activity on the aggregation of Josephin (Figure 2B, Table 1), whereby inhibition of Josephin aggregation is enhanced compared to the full-length at3(Q64) variant. Thus, an equimolar concentration of  $\alpha$ B-c completely suppresses Josephin aggregation but at3(Q64) still forms some ThT-reactive aggregates, even when  $\alpha$ B-c is present in a 2:1 molar excess. The reaction timecourse is too long to determine if the biphasic ThT pattern is altered at stoichiometric concentrations of  $\alpha$ B-c. The potent inhibitory effect of  $\alpha$ B-c on Josephin aggregation, combined with the data in Figure 1 showing that  $\alpha$ B-c does not inhibit polyQ-dependent aggregation, suggest that the slow ThT phase in the presence of high  $\alpha$ B-c concentrations is dominated by polyQ intermolecular interactions.

We analyzed the Josephin and at3(Q64) fibrils by TEM (Figure 3A). Josephin forms short curvilinear fibrils, 5-10 nm in width and 10-100 nm long, when incubated alone. However, no fibrils were observed in samples incubated with equimolar concentrations of  $\alpha$ B-c. Small, spherical aggregates characteristic of oligomeric  $\alpha$ B-c are observed (Figure 3A). After 7 h of incubation, curvilinear worm-like aggregates, 5-15 nm in width, consistent with the previously described morphology of stage one aggregates (21), were observed in samples of at3(Q64) incubated alone. In the presence of  $\alpha$ B-c, there were very few small aggregates present after 7 h. After 24 h, at3(Q64) was beginning to form larger fibrillar structures, ranging from 200-300 nm in length and 12-30 nm in width. These are consistent with the morphology of stage 2/endpoint fibrils. There were very few stage two fibrils present in the samples incubated with  $\alpha$ B-c.

**At3(Q64) forms SDS-insoluble fibrils at a rate modulated by increasing  $\alpha$ B-c concentrations.**

We tested the effect of  $\alpha$ B-c on the formation of SDS-resistant at3(Q64) fibrils using the filter trap assay. Analysis of the membranes using densitometry showed that  $\alpha$ B-c retards the formation of SDS-insoluble fibrils, as demonstrated by a two-fold increase in half-life for at3(Q64) aggregation occurring in the presence of substoichiometric concentrations of  $\alpha$ B-c (at3(Q64) 10:1  $\alpha$ B-c) (Figure 3B, Table 1). The inhibitory effect of  $\alpha$ B-c is more pronounced at stoichiometric concentrations where SDS-insoluble fibrils were still formed but at much lower amounts relative to at3(Q64) incubated alone. The equimolar data could not be fit to the classic nucleation-dependent model because aggregation is incomplete over the 300 h timescale.



---

**The presence of  $\alpha$ B-crystallin retains ataxin-3 in a monomeric conformation**

$\alpha$ B-c exists as a dynamic oligomeric complex (2). Depending on the target protein it can irreversibly or transiently bind to target proteins, thereby inhibiting aggregation. To determine the stability of the  $\alpha$ B-c:at3 association we performed size-exclusion chromatography (SEC) of Josephin and at3(Q64) incubated with equimolar concentrations of  $\alpha$ B-c (Figure 4).

Monomeric Josephin eluted from a Superose 6 column at approximately 18.9 ml. After 24 h incubation, Josephin formed insoluble aggregates that were removed by centrifugation prior to injection onto the column, therefore the monomeric protein peak eluting at 18.9 ml was diminished (Figure 4A).  $\alpha$ B-c displayed a broad peak centered around 14 ml, consistent with  $\alpha$ B-c being present as a polydisperse oligomer. In the presence of  $\alpha$ B-c, most of the Josephin is retained in a monomeric conformation at 49 h, with some recruited into the  $\alpha$ B-c complex, observed by a distinct shift and broadening of the  $\alpha$ B-c peak (Figure 4B).

At3(Q64) eluted as a single peak with a retention volume of approximately 17.8 ml. After 8 h, most of the at3(Q64) had aggregated, indicated by the presence of a large peak in the void volume (approximately 8 ml) of the Superose 6 column (Figure 4C). After 8 h in the presence of  $\alpha$ B-c most of the at3(Q64) is retained in a monomeric conformation, with little change in the elution profile after this time (Figure 4D). The SEC data suggest that  $\alpha$ B-c does not form a stable complex with at3/Josephin maintaining the proteins in a monomeric conformation.

**Addition of  $\alpha$ B-c affects Josephin/at3(Q64) fibril propagation**

The observation that  $\alpha$ B-c retains Josephin and at3(Q64) in a monomeric conformation suggests that  $\alpha$ B-c inhibits the conformational changes involved in nucleation of aggregation.  $\alpha$ B-c has also been shown to inhibit fibril propagation in a number of

systems (6, 12, 16, 29, 34). We tested the effect of  $\alpha$ B-c on Josephin and at3(Q64) fibril propagation by adding  $\alpha$ B-c at various intervals after the commencement of aggregation.

When  $\alpha$ B-c was added to isolated Josephin at 48 and 96 h there was no further increase in ThT intensity, suggesting that  $\alpha$ B-c prevents the further propagation of Josephin fibrils (Figure 5A). Similar behaviour is observed when  $\alpha$ B-c is added at timepoints along the aggregation pathway of  $\alpha$ -synuclein (6). When  $\alpha$ B-c was added to the at3(Q64) reactions at various intervals, there was a flattening of the curve for a period of time, then a second slower aggregation phase (Figure 5B). Because of the absence of this phase in the isolated Josephin reactions we predict that the second phase is due to polyQ-dependent aggregation. Interestingly, the length of the flattened period was longer when  $\alpha$ B-c was added at earlier timepoints in comparison to later timepoints, supporting the hypothesis that the rate at which poly(Q) interactions occurs is dependent on the initial formation of stage one Josephin-dependent fibrils.

### **$\alpha$ B-c interacts with an extensive region of Josephin**

The ability of  $\alpha$ B-c to inhibit at3 aggregation appears to be derived from inhibition of the Josephin-dependent stage (stage one) of the multidomain aggregation pathway. To provide insight into the mechanism of Josephin aggregation and chaperone inhibition of this process we used NMR spectroscopy to probe the molecular details of the Josephin:  $\alpha$ B-c interaction surface.

The NMR data show that  $\alpha$ B-c interacts with an extensive region of Josephin (Figure 6). A number of backbone amide cross-peaks in the  $^{15}\text{N}$ -HSQC spectrum of Josephin were significantly broadened upon titration of  $\alpha$ B-c up to a five-fold stoichiometric excess. The broadening is consistent with exchange between the free and

$\alpha$ B-c bound states of Josephin. Exchange broadening occurs when the timescale of the interaction is commensurate with the NMR  $^1\text{H}$  frequency (600MHz) which is consistent with the transient interaction mechanism observed by SEC (Figure 4). The Josephin residues perturbed to the largest degree were Q16, L19, N21, E26, S29, S35, I36, I77, W87, Q100, C114, V123, G127, F131, S135, L137, T138, L148, L155, Q156, E158, I162, Q176 and R182. These residues are confined to the globular subdomain of Josephin and include the active site residue C114.

Since chemical shift perturbations can arise through indirect mechanisms (rather than a direct interaction between  $\alpha$ B-c and Josephin), we used the paramagnetic relaxation agent Gd(DTPA-BMA) to independently confirm the interaction surface (35, 36). The presence of Gd(DTPA-BMA) in solution causes paramagnetic relaxation enhancement (PRE), broadening the protein resonances. By performing these  $^{15}\text{N}$ -HSQC experiments in the presence and absence of  $\alpha$ B-c we could determine the  $\alpha$ B-c interaction surface of Josephin, as the regions that interact with  $\alpha$ B-c were partially shielded from paramagnetic relaxation. Josephin residues that had >20 % stronger signals in the presence of Gd(DTPA-BMA) and a two-fold excess of  $\alpha$ B-c were S3, G11, L13, Q24, F28, A49, E50, G51, G52, T54, E56, D57, R59, T60, F61, S66, D71, G73, F75, S76, E90, N115, K128, N132, G139, L142, F151, G159, D168, E173, L178, M180 and I181. The identity of these residues correlates well with those broadened in the  $\alpha$ B-c-bound state of Josephin (see above, Figure 6). When mapped to the Josephin structure all of the residues that appeared to be most shielded from the paramagnetic probe by the presence of  $\alpha$ B-c were found to be surface exposed, therefore validating these regions as the  $\alpha$ B-c binding interface.

---

**Discussion****sHsps do not inhibit polyQ aggregation**

sHsps have a critical role in cellular homeostasis. Their upregulation in several neurodegenerative disease states suggests that they can provide a barrier to the aggregation of partially unfolded proteins that are central to many neurodegenerative diseases. We investigated the effect of  $\alpha$ B-c on the aggregation of a polyQ-containing model protein, finding that  $\alpha$ B-c does not affect the formation of SDS-resistant polyQ fibrils when fibril formation is entirely polyQ-dependent.

These data show that  $\alpha$ B-c does not recognize misfolded polyQ target proteins (Fig. 1). Typically sHsps (along with many other molecular chaperones) interact with exposed hydrophobic patches on target proteins. Glutamine is a hydrophilic amino acid and the polyQ region is unfolded in the native state (37). As a result, polyQ regions do not undergo an increase in hydrophobicity during their early aggregation stages. In a manner akin to our findings with sHsps, molecular chaperones from other classes, e.g. Hsp70 and the chaperonin TRiC, which recognize surface hydrophobic patches of amyloidogenic intermediates, only weakly associate with polyQ proteins (30, 38). Therefore, the inherent properties of the polyQ tract may lead to its evasion by some chaperone protective mechanisms.

 **$\alpha$ B-c inhibits the Josephin-dependent stage of ataxin-3 aggregation**

In the physiological context, pathological polyQ tracts are located within specific host proteins and properties of the host protein are known to modulate aggregation and cellular toxicity (39, 40). Although molecular chaperones do not alter the misfolding of pure polyQ, genetic screens in *Drosophila* and *C. elegans* have identified several chaperones that modulate toxicity induced by ataxin-3, ataxin-1 and Htt exon1, proteins

that contain well-structured domain(s) coupled to a polyQ repeat (10, 41). As a result, it is hypothesized that non-polyQ aggregation mechanisms play a role in pathogenesis. A previous study from our group has described a two-stage aggregation mechanism for pathological variants of ataxin-3 (21). Our data show that  $\alpha$ B-c inhibits nucleation of the Josephin-dependent stage of at-3 aggregation, subsequently having an indirect effect on the rate of poly(Q) intermolecular interactions.

$\alpha$ B-c had a more potent inhibitory effect on the aggregation of Josephin alone compared with at3(Q64). The Josephin aggregation stage is enhanced when this domain is part of at3(Q64), suggesting that the presence of the polyQ C-terminal tail increases the rate of Josephin aggregation. The presence of the C-terminal polyQ tail leads to a slight decrease in the thermodynamic stability of Josephin (42), and we speculate that the tail alters the dynamic properties of the core domain enhancing the rate of conformational rearrangements on the aggregation pathway. The observation that  $\alpha$ B-c has a greater inhibitory effect on the aggregation of isolated Josephin can be rationalized by a previous study demonstrating that larger target proteins require more  $\alpha$ B-c (on a molar basis) for complete suppression of aggregation (43). The altered potency is also consistent with data showing that  $\alpha$ B-c has a preference for target proteins that aggregate more slowly (44). This relationship has been described for an amorphously aggregating system (44), however this is the first report suggesting that the aggregation kinetics of a fibril-forming protein are a determinant for  $\alpha$ B-c chaperone efficiency.

For the full-length protein, the interaction of  $\alpha$ B-c with Josephin significantly slows the subsequent poly(Q)-dependent aggregation (Figure 1, Table 1). These data suggest that Josephin self interactions lower the kinetic barrier to subsequent poly(Q)-dependent aggregation. The multidomain aggregation mechanisms described for ataxin-

3 and Htt exon 1 suggest that a pool of small stage one species can enhance the rate of polyQ-dependent aggregation by increasing the local polyQ concentration (21, 22, 45). This is consistent with the data in Figure 5B, in which  $\alpha$ B-c was added at various stages of the aggregation reaction demonstrating that the formation of small Josephin fibrils is necessary for efficient formation of poly(Q)-dependent SDS-insoluble fibrils.

The  $\alpha$ B-c binding interface of Josephin is partially coincident with the Ubiquitin and HHR23B interacting surfaces (46-48). It would therefore be interesting to determine the effect of Ubiquitin and HHR23B on Josephin-dependent aggregation. In the crowded cellular environment, the cumulative effect of these interaction networks could significantly affect the intracellular aggregation mechanism of Josephin and Josephin-containing proteins.

The significant effect of  $\alpha$ B-c on ataxin-3 aggregation is consistent with a *Drosophila* model in which overexpression of  $\alpha$ B-c decreased SCA3-induced toxicity (10), suggesting that inhibition of aggregation by chaperones is relevant *in vivo*. Another *Drosophila* study demonstrated that chaperone-induced retention of at3 in a monomeric conformation suppresses toxicity (49). Conversely,  $\alpha$ B-c overexpression had no effect on the HD phenotype (27). Htt exon 1 has been proposed to have a similar two-stage aggregation mechanism compared to ataxin-3 (22). The alternative effects of  $\alpha$ B-c in the ataxin-3 and Htt systems demonstrate how specific differences in multidomain misfolding pathways could lead to variable cellular responses.

### **$\alpha$ B-c transiently interacts with monomeric Josephin**

The aggregation data clearly show that  $\alpha$ B-c suppresses aggregation by associating with species on the Josephin aggregation pathway. Our SEC and NMR data show that inhibition mechanistically occurs by transient association of  $\alpha$ B-c with a defined region

on the Josephin surface. The SEC data show that the proteins are retained in a monomeric conformation in the presence of  $\alpha$ B-c (Figure 4), and the NMR data confirm this as specific cross-peaks in the  $^{15}\text{N}$ -HSQC spectrum of monomeric Josephin exchange-broadened, implying that Josephin is in intermediate exchange in the presence of  $\alpha$ B-c. These data suggest that the rate at which  $\alpha$ B-c interacts with Josephin is more rapid than the rate of misfolding. The NMR data indicate that this is a low affinity interaction, therefore, during this interaction, most of the Josephin is in a monomeric uncomplexed state. Together with the data indicating that  $\alpha$ B-c addition prevents the further propagation of Josephin-dependent fibrils (Figure 5), these data are consistent with a mechanism whereby  $\alpha$ B-c binds and sequesters the free pool of monomeric proteins thereby preventing fibril growth (Figure 7). As Josephin is a dynamic domain (42) and populates a number of native conformations (Saunders HM and Bottomley SP *et al*, unpublished data), evaluating the effect of  $\alpha$ B-c on the equilibrium between these states will be addressed in further work.

The transient interaction between  $\alpha$ B-c and monomeric Josephin, encompassing a large region of the target protein, is also observed, as monitored by NMR spectroscopy, in the interaction of  $\alpha$ B-c with the fibril-forming proteins,  $\alpha$ -synuclein (29),  $\kappa$ -casein (6) and  $\beta_2$ -microglobulin (Esposito G and Carver JA, unpublished data).

Using peptide screening our group has mapped the Josephin fibril core, finding that the region comprising L91-S96 and T146-G167 has greatest fibrillogenic propensity (Ellisdon AM, Saunders HM and Bottomley SP, unpublished data). These segments localize to a contiguous structural region of Josephin comprising the  $\alpha$ -helix  $\alpha 5$ , and  $\beta$ -strands  $\beta 1$  and  $\beta 5$  when mapped onto the NMR solution structure determined by Pastore and colleagues (48). This region is well structured and partially buried within the protein, suggesting that significant conformational rearrangements are required for

aggregation to occur. The NMR data show that  $\alpha$ B-c interacts with a broad region of Josephin, and some of the residues within this region are in  $\alpha$ 5 and  $\beta$ 5, therefore overlaying the proposed fibrillogenic core. Residues in this region affected by  $\alpha$ B-c binding are not surface exposed, and are likely affected by conformational perturbations that occur upon interaction of  $\alpha$ B-c with surface residues. The  $\alpha$ B-c binding interface of Josephin in fact maps largely to the face opposing the most fibrillogenic region. We speculate that binding of  $\alpha$ B-c inhibits partial unfolding that is required for exposure of the fibrillogenic core residues involved in nucleation of the Josephin aggregation stage.

## Conclusions

Our data demonstrate that inhibition of the non-polyQ aggregation stage by a molecular chaperone can significantly impede the cumulative protein aggregation process involved in polyQ-related diseases. These data strongly support the two-stage model as a plausible scheme for ataxin-3 aggregation *in vitro*. Therapeutic upregulation of chaperone activity may be beneficial in treating SCA3 and other polyQ diseases that have a multidomain aggregation mechanism.



---

**Experimental Procedures****Preparation of proteins**

SpA-cQ52 was expressed and purified as previously described (28). Ataxin-3 variants were expressed and purified as described previously (50).  $\alpha$ B-c was expressed and purified as described in (51).

**Protein aggregation assays**

Aggregation assays were performed for SpA-cQ52 and ataxin-3 variants at a protein concentration of 20  $\mu$ M in Tris-buffered saline (TBS) (100 mM Tris (pH 7.4), 80 mM NaCl) containing 2 mM PMSF, 5 mM EDTA and 15 mM  $\beta$ -mercaptoethanol. Variable concentrations of  $\alpha$ B-c were added to the samples, ranging from 2  $\mu$ M to 40  $\mu$ M. Samples were incubated at 37 °C and stored in airtight containers to avoid evaporation.

**ThT measurements**

ThT fluorescence measurements were performed continuously using a BMG Laboratories FLUOstar Optima fluorescence platereader using excitation and emission wavelengths of 450 nm and 490 nm respectively. Reactions were carried out as described above, with the addition of 20  $\mu$ M ThT.

**Membrane Filter trap assay**

The membrane filter-trap assay was performed as previously described (21).

---

**Transmission electron microscopy**

Samples were adsorbed onto carbon grids and negatively stained with 1 % w/v aqueous uranyl acetate (pH 8). Samples were then analysed by transmission electron microscopy on a Hitachi H7500 electron microscope using an acceleration voltage of 80 kV.

**Circular Dichroism**

Samples in TBS were analysed by circular dichroism (CD) on a Jasco-810 spectropolarimeter at 25 °C using a 0.01 cm pathlength cuvette. Spectra were recorded from 195 to 260 nm using a scan speed of 50 nm/min.

**Size exclusion chromatography**

Size exclusion chromatography (SEC) of at3(Q64) and Josephin in the presence and absence of equimolar concentrations of  $\alpha$ B-c was performed on an Akta-FPLC equipped with a Superose 6 10/300 GL column (Pharmacia). Prior to injection onto the column samples were centrifuged for 5 minutes at 13000 rpm to remove insoluble aggregated material. The absorbance at 214 nm was measured as proteins were eluted in TBS buffer at a constant flow rate of 0.5 ml/min.

**NMR spectroscopy**

Josephin samples for NMR contained  $^{15}\text{N}$ -labelled Josephin (0.3 mM) or  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labelled Josephin (0.5 mM) in 50 mM ammonium acetate, 50 mM ammonium carbonate pH 7.4. All spectra were recorded at 25 °C on a Varian Inova 600 NMR spectrometer equipped with a cryoprobe, operating at a  $^1\text{H}$  frequency of 600 MHz.

To investigate the  $\alpha$ B-c binding site on Josephin by chemical shift perturbation  $^{15}\text{N}$ -HSQC spectra were acquired in the presence and absence of unlabelled  $\alpha$ B-c at

$\alpha$ B-c:Josephin molar ratios of 0, 0.1, 0.25, 0.6, 2 and 5. The spectral dimensions were 2048 (t<sub>2</sub>) x 160 (t<sub>1</sub>) increments with 32 scans per increment and a recycle delay of 1.2 s.

The surface-exposed residues involved in the interaction were probed using paramagnetic relaxation enhancement (PRE). A gadolinium-based paramagnetic relaxation agent, Gd-diethylenetriamine pentaacetic acid-bismethylamide (Gd(DTPA-BMA)) (36) at a concentration of 5 mM was added to a 0.3 mM <sup>15</sup>N-labelled Josephin sample. <sup>15</sup>N-HSQC spectra were recorded in the absence and presence of 0.6 mM  $\alpha$ B-c.

For assignment purposes 3D HNCO, HNCA and CBCA(CO)NH spectra (52) were acquired using a 0.3 mM <sup>13</sup>C, <sup>15</sup>N-Josephin sample. All data were processed using nmrPipe (53) with sine squared window functions, zero filling and linear prediction in the indirect dimensions. Spectra were visualised and analysed using Sparky (Goddard and Keller, UCSF).

---

**References**

1. Haslbeck M, Franzmann T, Weinfurter D, & Buchner J (2005) Some like it hot: the structure and function of small heat-shock proteins. (Translated from eng) *Nat Struct Mol Biol* 12(10):842-846 (in eng).
2. Ecroyd H & Carver JA (2009) Crystallin proteins and amyloid fibrils. (Translated from eng) *Cell Mol Life Sci* 66(1):62-81 (in eng).
3. Horwitz J (2003) Alpha-crystallin. (Translated from eng) *Experimental eye research* 76(2):145-153 (in eng).
4. Iwaki T, *et al.* (1992) Accumulation of alpha B-crystallin in central nervous system glia and neurons in pathologic conditions. (Translated from eng) *The American journal of pathology* 140(2):345-356 (in eng).
5. Das KP & Surewicz WK (1995) Temperature-induced exposure of hydrophobic surfaces and its effect on the chaperone activity of alpha-crystallin. (Translated from eng) *FEBS letters* 369(2-3):321-325 (in eng).
6. Rekas A, Jankova L, Thorn DC, Cappai R, & Carver JA (2007) Monitoring the prevention of amyloid fibril formation by alpha-crystallin. Temperature dependence and the nature of the aggregating species. (Translated from eng) *The FEBS journal* 274(24):6290-6304 (in eng).
7. Coi A, *et al.* (2008) Structural perturbation of alphaB-crystallin by zinc and temperature related to its chaperone-like activity. (Translated from eng) *International journal of biological macromolecules* 42(3):229-234 (in eng).
8. Lowe J, *et al.* (1992) alpha B crystallin expression in non-lenticular tissues and selective presence in ubiquitinated inclusion bodies in human disease. (Translated from eng) *The Journal of pathology* 166(1):61-68 (in eng).
9. Renkawek K, *et al.* (1992) alpha B-crystallin is present in reactive glia in Creutzfeldt-Jakob disease. (Translated from eng) *Acta neuropathologica* 83(3):324-327 (in eng).
10. Bilen J & Bonini NM (2007) Genome-wide screen for modifiers of ataxin-3 neurodegeneration in Drosophila. (Translated from eng) *PLoS Genet* 3(10):1950-1964 (in eng).
11. Wilhelmus MM, *et al.* (2006) Small heat shock proteins inhibit amyloid-beta protein aggregation and cerebrovascular amyloid-beta protein toxicity. (Translated from eng) *Brain Res* 1089(1):67-78 (in eng).
12. Hatters DM, Lindner RA, Carver JA, & Howlett GJ (2001) The molecular chaperone, alpha-crystallin, inhibits amyloid formation by apolipoprotein C-II. (Translated from eng) *The Journal of biological chemistry* 276(36):33755-33761 (in eng).
13. Stege GJ, *et al.* (1999) The molecular chaperone alphaB-crystallin enhances amyloid beta neurotoxicity. (Translated from eng) *Biochemical and biophysical research communications* 262(1):152-156 (in eng).
14. Carver JA, *et al.* (2002) The interaction of the molecular chaperone alpha-crystallin with unfolding alpha-lactalbumin: a structural and kinetic spectroscopic study. (Translated from eng) *Journal of molecular biology* 318(3):815-827 (in eng).
15. Devlin GL, Carver JA, & Bottomley SP (2003) The selective inhibition of serpin aggregation by the molecular chaperone, alpha-crystallin, indicates a nucleation-dependent specificity. (Translated from eng) *The Journal of biological chemistry* 278(49):48644-48650 (in eng).

16. Raman B, *et al.* (2005) AlphaB-crystallin, a small heat-shock protein, prevents the amyloid fibril growth of an amyloid beta-peptide and beta2-microglobulin. (Translated from eng) *The Biochemical journal* 392(Pt 3):573-581 (in eng).
17. Ecroyd H, *et al.* (2007) Mimicking phosphorylation of alphaB-crystallin affects its chaperone activity. (Translated from eng) *The Biochemical journal* 401(1):129-141 (in eng).
18. Chen S, Ferrone FA, & Wetzel R (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. (Translated from eng) *Proc Natl Acad Sci U S A* 99(18):11884-11889 (in eng).
19. Ignatova Z & Gierasch LM (2006) Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein. (Translated from eng) *J Biol Chem* 281(18):12959-12967 (in eng).
20. Ellisdon AM, Pearce MC, & Bottomley SP (2007) Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. (Translated from eng) *J Mol Biol* 368(2):595-605 (in eng).
21. Ellisdon AM, Thomas B, & Bottomley SP (2006) The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. (Translated from eng) *The Journal of biological chemistry* 281(25):16888-16896 (in eng).
22. Thakur AK, *et al.* (2009) Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. (Translated from eng) *Nat Struct Mol Biol* 16(4):380-389 (in eng).
23. Bulone D, Masino L, Thomas DJ, San Biagio PL, & Pastore A (2006) The interplay between PolyQ and protein context delays aggregation by forming a reservoir of protofibrils. (Translated from eng) *PLoS ONE* 1:e111 (in eng).
24. de Chiara C, *et al.* (2005) The AXH domain adopts alternative folds the solution structure of HBP1 AXH. (Translated from eng) *Structure* 13(5):743-753 (in eng).
25. Wytenbach A, *et al.* (2002) Heat shock protein 27 prevents cellular polyglutamine toxicity and suppresses the increase of reactive oxygen species caused by huntingtin. (Translated from eng) *Hum Mol Genet* 11(9):1137-1151 (in eng).
26. Carra S, Brunsting JF, Lambert H, Landry J, & Kampinga HH (2009) HspB8 participates in protein quality control by a non-chaperone-like mechanism that requires eIF2(alpha) phosphorylation. (Translated from eng) *The Journal of biological chemistry* 284(9):5523-5532 (in eng).
27. Carra S, Sivilotti M, Chavez Zobel AT, Lambert H, & Landry J (2005) HspB8, a small heat shock protein mutated in human neuromuscular disorders, has in vivo chaperone activity in cultured cells. (Translated from eng) *Hum Mol Genet* 14(12):1659-1669 (in eng).
28. Robertson AL, *et al.* (2008) The structural impact of a polyglutamine tract is location-dependent. (Translated from eng) *Biophys J* 95(12):5922-5930 (in eng).
29. Rekas A, *et al.* (2004) Interaction of the molecular chaperone alphaB-crystallin with alpha-synuclein: effects on amyloid fibril formation and chaperone activity. (Translated from eng) *Journal of molecular biology* 340(5):1167-1183 (in eng).
30. Muchowski PJ, *et al.* (2000) Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. (Translated from eng) *Proceedings of the National Academy of Sciences of the United States of America* 97(14):7841-7846 (in eng).

31. Scherzinger E, *et al.* (1997) Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. (Translated from eng) *Cell* 90(3):549-558 (in eng).
32. Scherzinger E, *et al.* (1999) Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. (Translated from eng) *Proceedings of the National Academy of Sciences of the United States of America* 96(8):4604-4609 (in eng).
33. Masino L, *et al.* (2004) Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3. (Translated from eng) *Journal of molecular biology* 344(4):1021-1035 (in eng).
34. Knowles TP, *et al.* (2007) Kinetics and thermodynamics of amyloid formation from direct measurements of fluctuations in fibril mass. (Translated from eng) *Proceedings of the National Academy of Sciences of the United States of America* 104(24):10016-10021 (in eng).
35. Liepinsh E, *et al.* (2001) Thioredoxin fold as homodimerization module in the putative chaperone ERp29: NMR structures of the domains and experimental model of the 51 kDa dimer. (Translated from eng) *Structure* 9(6):457-471 (in eng).
36. Pintacuda G & Otting G (2002) Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. (Translated from eng) *J Am Chem Soc* 124(3):372-373 (in eng).
37. Klein FA, *et al.* (2007) Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. (Translated from eng) *J Mol Biol* 371(1):235-244 (in eng).
38. Behrends C, *et al.* (2006) Chaperonin TRiC promotes the assembly of polyQ expansion proteins into nontoxic oligomers. (Translated from eng) *Mol Cell* 23(6):887-897 (in eng).
39. Nozaki K, Onodera O, Takano H, & Tsuji S (2001) Amino acid sequences flanking polyglutamine stretches influence their potential for aggregate formation. (Translated from eng) *Neuroreport* 12(15):3357-3364 (in eng).
40. Duennwald ML, Jagadish S, Muchowski PJ, & Lindquist S (2006) Flanking sequences profoundly alter polyglutamine toxicity in yeast. (Translated from eng) *Proc Natl Acad Sci U S A* 103(29):11045-11050 (in eng).
41. Branco J, *et al.* (2008) Comparative analysis of genetic modifiers in Drosophila points to common and distinct mechanisms of pathogenesis among polyglutamine diseases. (Translated from eng) *Hum Mol Genet* 17(3):376-390 (in eng).
42. Chow MK, Mackay JP, Whisstock JC, Scanlon MJ, & Bottomley SP (2004) Structural and functional analysis of the Josephin domain of the polyglutamine protein ataxin-3. (Translated from eng) *Biochem Biophys Res Commun* 322(2):387-394 (in eng).
43. Lindner RA, Kapur A, Mariani M, Titmuss SJ, & Carver JA (1998) Structural alterations of alpha-crystallin during its chaperone action. (Translated from eng) *European journal of biochemistry / FEBS* 258(1):170-183 (in eng).
44. Lindner RA, Kapur A, & Carver JA (1997) The interaction of the molecular chaperone, alpha-crystallin, with molten globule states of bovine alpha-lactalbumin. (Translated from eng) *The Journal of biological chemistry* 272(44):27722-27729 (in eng).

45. Ignatova Z, Thakur AK, Wetzel R, & Gierasch LM (2007) In-cell aggregation of a polyglutamine-containing chimera is a multistep process initiated by the flanking sequence. (Translated from eng) *J Biol Chem* 282(50):36736-36743 (in eng).
46. Nicastro G, Habeck M, Masino L, Svergun DI, & Pastore A (2006) Structure validation of the Josephin domain of ataxin-3: conclusive evidence for an open conformation. (Translated from eng) *J Biomol NMR* 36(4):267-277 (in eng).
47. Nicastro G, *et al.* (2009) The josephin domain of ataxin-3 contains two distinct ubiquitin binding sites. (Translated from Eng) *Biopolymers* (in Eng).
48. Nicastro G, *et al.* (2005) The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. (Translated from eng) *Proceedings of the National Academy of Sciences of the United States of America* 102(30):10493-10498 (in eng).
49. Chan HY, Warrick JM, Gray-Board GL, Paulson HL, & Bonini NM (2000) Mechanisms of chaperone suppression of polyglutamine disease: selectivity, synergy and modulation of protein solubility in *Drosophila*. (Translated from eng) *Hum Mol Genet* 9(19):2811-2820 (in eng).
50. Chow MK, Ellisdon AM, Cabrita LD, & Bottomley SP (2004) Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. (Translated from eng) *The Journal of biological chemistry* 279(46):47643-47651 (in eng).
51. Horwitz J, Huang QL, Ding L, & Bova MP (1998) Lens alpha-crystallin: chaperone-like properties. (Translated from eng) *Methods Enzymol* 290:365-383 (in eng).
52. Sattler M, Schleucher J, & Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* 34(2):93-158.
53. Delaglio F, *et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. (Translated from eng) *J Biomol NMR* 6(3):277-293 (in eng).

**Table 1:** Half-lives (in hours) of Ataxin-3(Q64) aggregation followed by ThT and SDS-insolubility in the presence and absence of  $\alpha$ B-c.

|                                 | ThT            |                | SDS-insolubility |
|---------------------------------|----------------|----------------|------------------|
|                                 | Stage 1        | Stage 2        |                  |
| Ataxin-3(Q64)                   | $9.3 \pm 0.2$  | $31.7 \pm 0.9$ | $25.5 \pm 1.1$   |
| Ataxin-3(Q64) 10:1 $\alpha$ B-c | $19.7 \pm 0.6$ | $49 \pm 4$     | $45.1 \pm 1.3$   |
| Ataxin-3(Q64) 1:1 $\alpha$ B-c  | #              | #              | #                |
| Josephin                        | $59.1 \pm 3$   | *              | *                |
| Josephin 10:1 $\alpha$ B-c      | $174 \pm 0.3$  | *              | *                |
| Josephin 1:1 $\alpha$ B-c       | -              | *              | *                |

\* Josephin does not undergo stage 2 of the ataxin-3 aggregation pathway and therefore does not form SDS-insoluble fibrils.

# Aggregation incomplete after 300h therefore data could not be fitted.

- An equimolar amount of  $\alpha$ B-c completely inhibited Josephin aggregation over 300h.



**Figure Legends**

**Figure 1: The effect of  $\alpha$ B-c on poly(Q) aggregation.** *A.* SpA-cQ52 aggregation in the presence of increasing concentrations of  $\alpha$ B-c monitored by ThT fluorescence. *B.* Analysis of the SDS-solubility of endpoint aggregates (100h) in the presence and absence of  $\alpha$ B-c. *C.* and *D.* TEM analysis of SpA-cQ52 aggregates in the absence (*C.*) and presence of  $\alpha$ B-c. Ratios of SpA-cQ52: $\alpha$ B-c are given on a molar basis.

**Figure 2: The effect of  $\alpha$ B-c on at3 aggregation.** *A.* At3(Q64) and *B.* Josephin aggregation in the presence of increasing concentrations of  $\alpha$ B-c monitored by ThT-fluorescence. Ratios of at3 variant: $\alpha$ B-c are given on a molar basis.

**Figure 3: The effect of  $\alpha$ B-c on the formation of at3(Q64) SDS-insoluble fibrils.** *A.* TEM analysis of Josephin and ataxin-3 fibrils. *B.* The poly(Q)-dependent stage of At3(Q64) aggregation was monitored by detection of SDS-insoluble fibrils using the membrane filter trap assay. Membranes were analysed by densitometry and data normalized.

**Figure 4: At3 aggregation in the presence and absence of  $\alpha$ B-c monitored by SEC.** Samples of Josephin and At3(Q64) incubated in the presence and absence of  $\alpha$ B-c were removed at indicated timepoints and analysed by SEC using a Superose 6 column. In the samples containing  $\alpha$ B-c, equimolar amounts of substrate:  $\alpha$ B-c were used.

**Figure 5: The effect of  $\alpha$ B-c addition after at3 aggregation commencement.**  $\alpha$ B-c was added to *A.* Josephin and *B.* at3(Q64) aggregation reactions at the indicated timepoints as the reactions were monitored by ThT fluorescence.

**Figure 6:  $\alpha$ B-c interacts with specific regions of Josephin.** *A.* *B.* The residues perturbed upon addition of  $\alpha$ B-c were mapped onto the Josephin domain (PDB ID: 1YZB). Residues in intermediate exchange in the presence of  $\alpha$ B-c (red), residues shielded from broadening by Gd(DTPA-BMA) in the presence of  $\alpha$ B-c (yellow), and orange shows the overlapping residues. *A.* and *B.* show opposing faces of the Josephin structure. *C.* *D.* The electrostatic surface of Josephin. Blue represents positive charge and red indicates negatively-charged regions. *C.* and *D.* show opposing faces of the Josephin structure.

**Figure 7: A schematic model describing the effect of  $\alpha$ B-c on at3 aggregation.** Multimeric  $\alpha$ B-c (a sHsp) transiently binds to at3 molecules in which Josephin is in an aggregation-prone conformation ( $J_{agg}$ ). The J:sHsp complex is unstable, and monomeric Josephin is released from the sHsp.

**Figures**

Figure 1

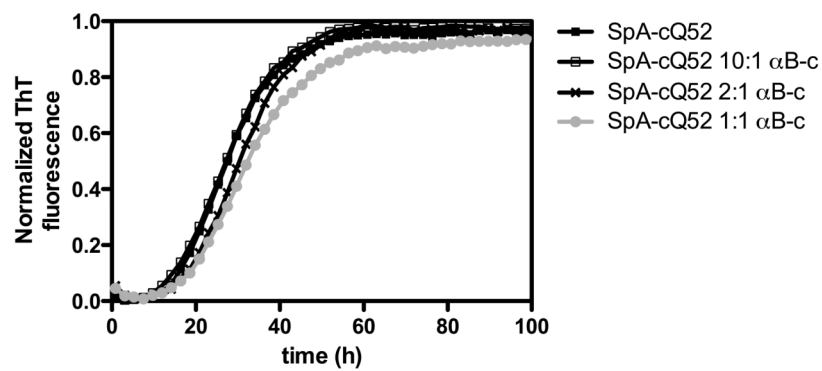
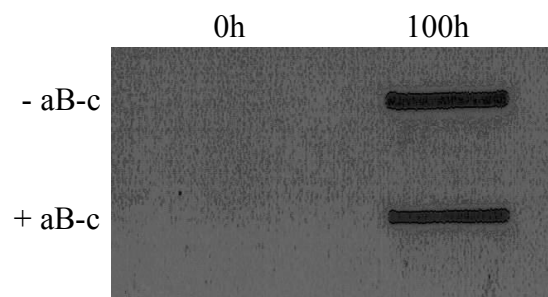
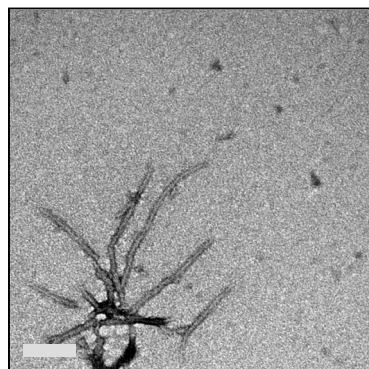
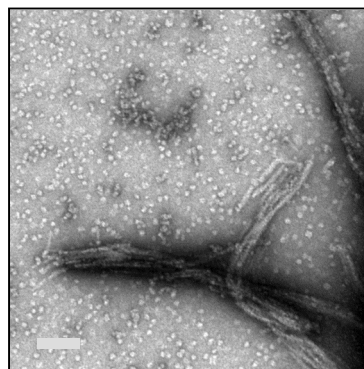
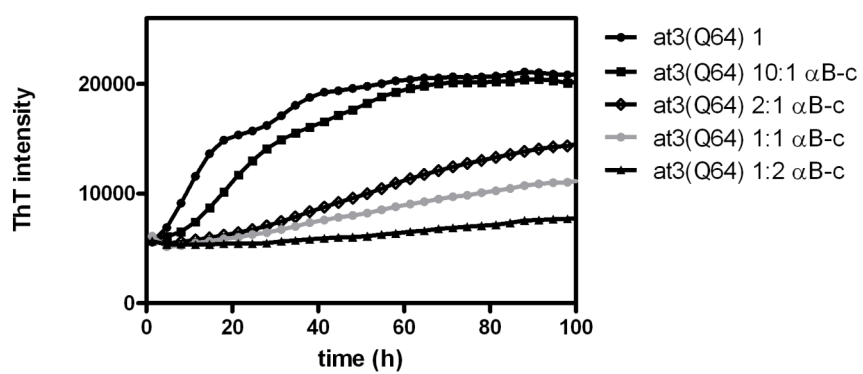
**A****B****C****D**

Figure 2

A



B

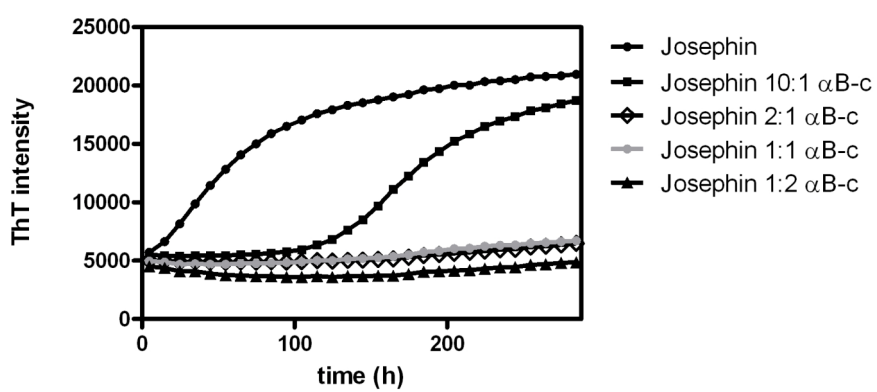
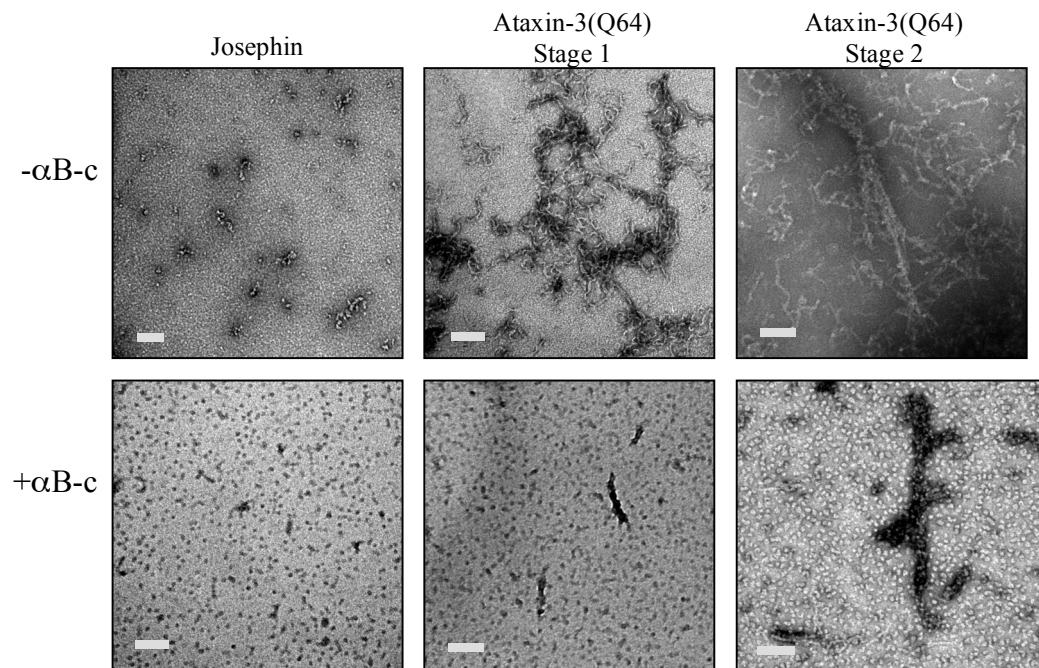


Figure 3

A.



B.

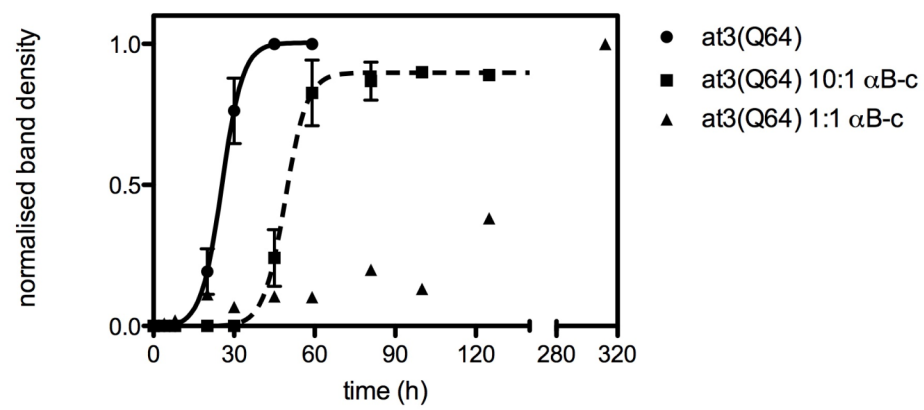


Figure 4

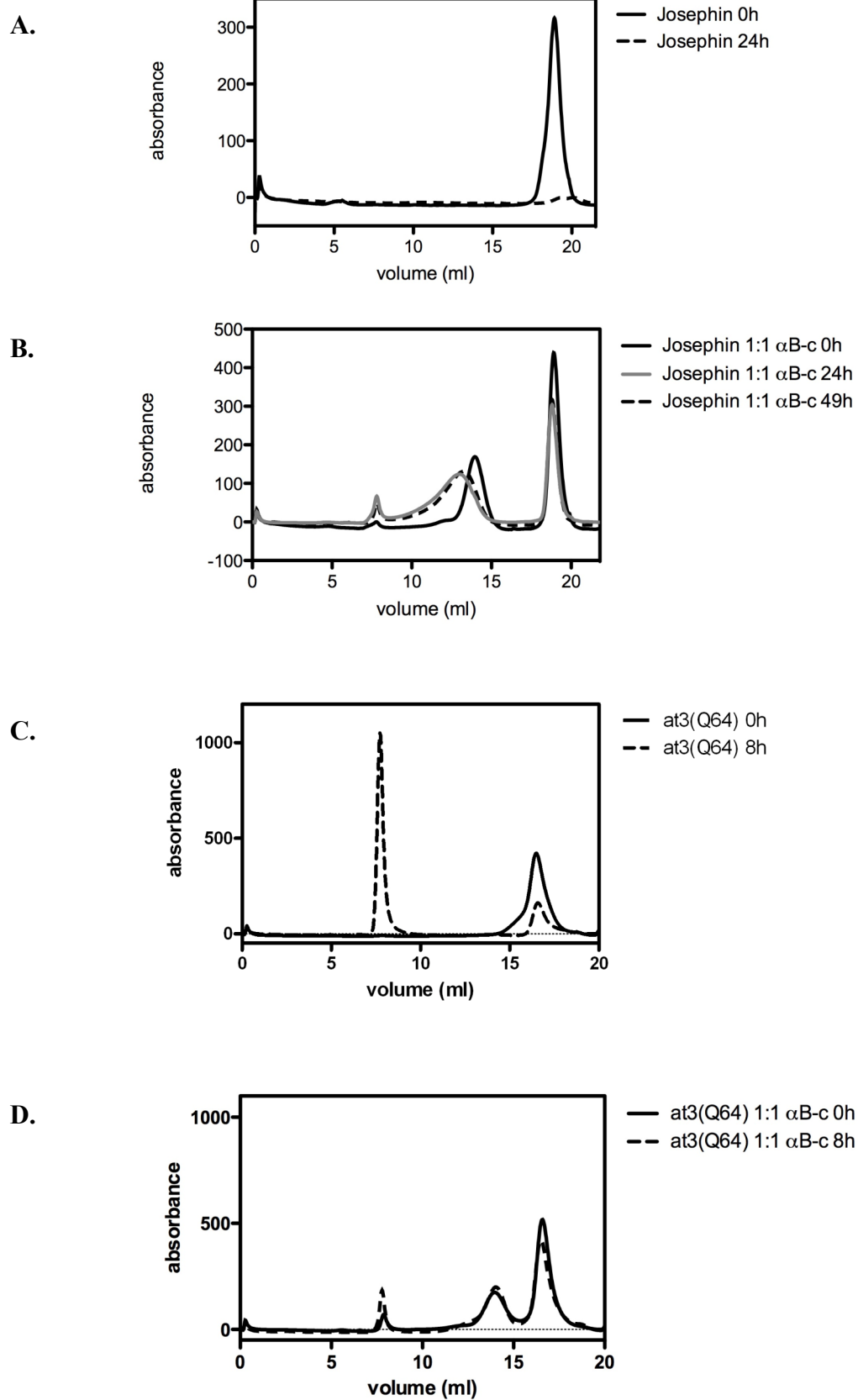


Figure 5

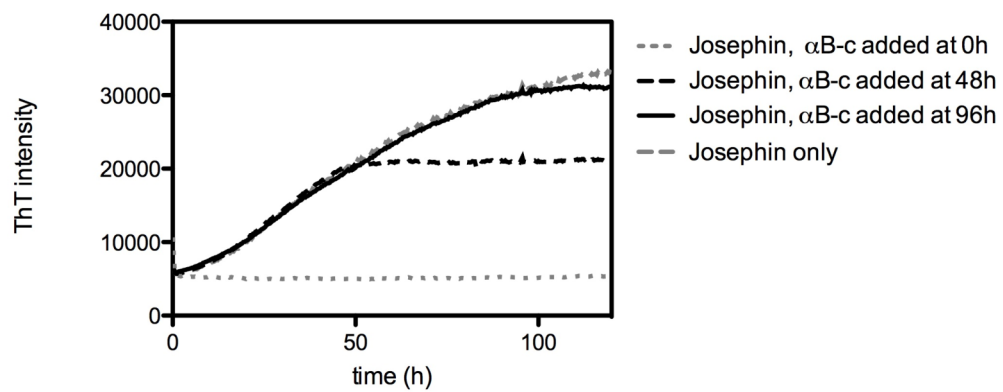
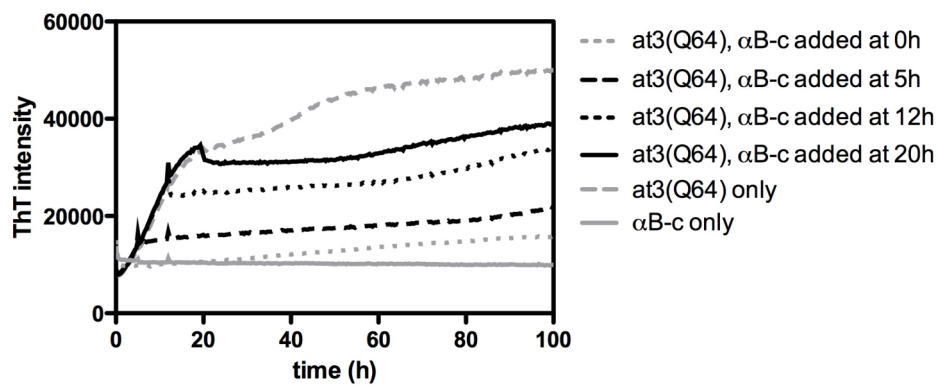
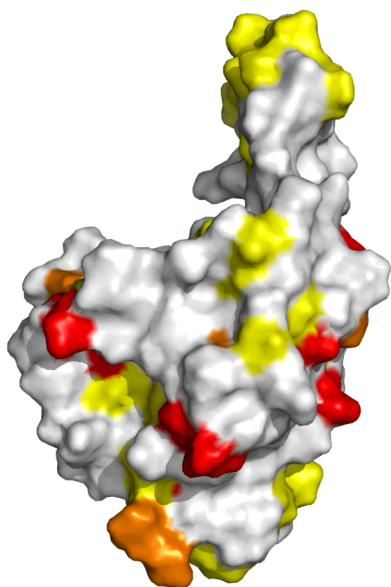
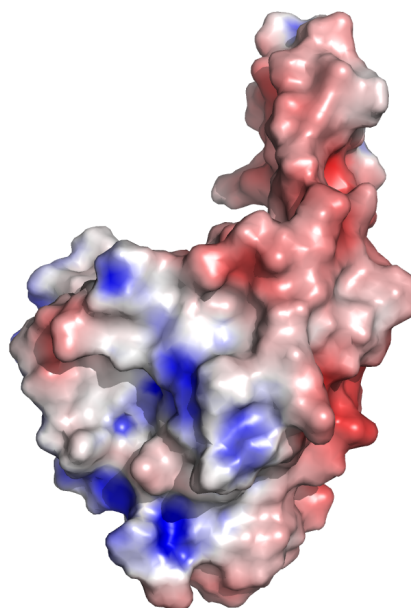
**A****B**

Figure 6

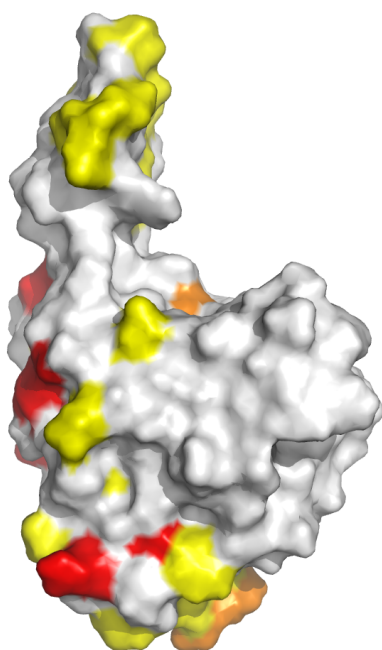
A.



C.



B.



D.

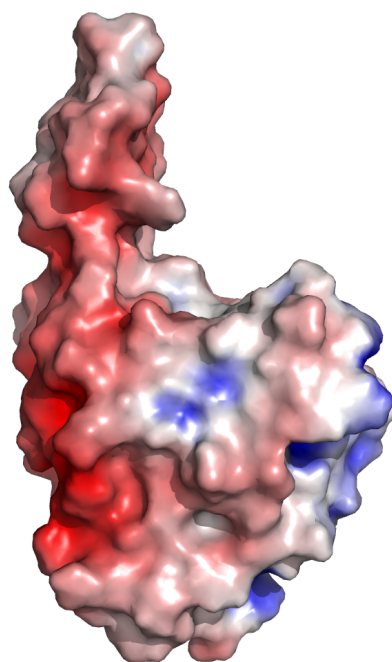
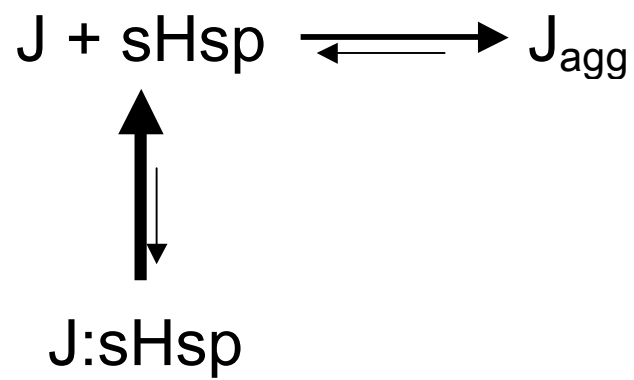


Figure 7





---

## **Chapter 7**

### **General Discussion**

---

## **7.1 Pathways to PolyQ Aggregation**

Whilst the formation of amyloid-like aggregates is a common mechanism leading to over 40 human diseases, and amyloid formation is thought to be a generic property of the polypeptide chain, properties of individual proteins lead to alternative misfolding pathways. There has been significant progress in the amyloid field since the outset of this project. Only recently, high-resolution structural descriptions of amyloid fibrils have become available (16, 17). Advancing techniques have been capable of mapping protein regions involved in the amyloid core (207-210), and further studies have highlighted that protein restructuring during aggregation can be global or localized (42, 44, 50). Furthermore, a number of experimental and theoretical studies have shown that regions flanking amyloid-prone sequences can modulate aggregation (128, 129, 211). In this thesis, the molecular mechanisms involved in polyQ aggregation and the role of the flanking protein regions in modulating the aggregation mechanism were examined. The results of this thesis have provided important insight into the mechanism of polyQ aggregation and have shown that the propensity for polyQ aggregation is determined by a complex interplay between the polyQ region, host protein factors and the cellular environment.

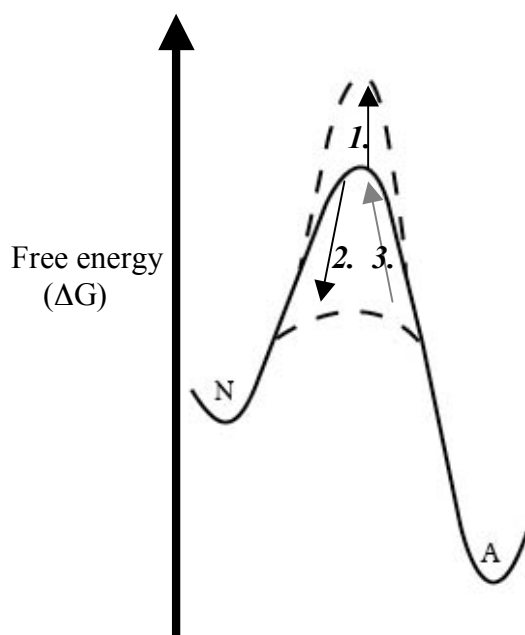
### **7.11 PolyQ Self-association**

It is widely considered that pathological length polyQ tracts lead to a toxic gain-of-function (92, 93). Within the disease proteins, the polyQ is located within a host protein. A number of studies have investigated the mechanisms by which pure polyQ peptides self-associate (63, 116), however, the precise molecular mechanisms for polyQ aggregation within a protein context have been limited by the inability to obtain recombinant versions of the disease proteins. Model protein systems studied thus far

have given varied results, and this may be because the environment of the polyQ has been different in each case, and recent evidence indicates that the structure of the polyQ tract is influenced by protein context (117, 122, 212). Given this, an important consideration in the design of a model protein is the location of the polyQ within the host protein. The first part of this thesis examined the molecular mechanism of polyQ aggregation within the context of a model protein, recapitulating the flexible, untethered environment of the polyQ tract observed in the disease proteins (Chapter 4). Our data clearly show that the stability and structure of a flanking domain are unaffected by polyQ repeat length, indicating that in the monomeric proteins, the polyQ region and flanking domain can behave independently. Our study confirmed previous work that the favoured conformation for monomeric polyQ is random coil, regardless of repeat length (122). Together these data suggest that local polyQ conformational fluctuations around the low energy native state (Figure 1.1), occurring at a rate faster than that of un/folding are the initial trigger for aggregation. In order for aggregation to occur, the polyQ region must overcome a kinetic barrier, which is lowered as repeat length increases (Figure 7.1).

### **7.12 Influence of Flanking Domains on PolyQ Aggregation**

Concurrent with the progress of this study, our group has examined the structure of the flanking domain in the context of the fibril (Thomas B *et al*, unpublished data, Appendix 1). Fourier transform infrared analyses suggest that the flanking domain retains its folded conformation within the fibril context. From these studies combined we describe a model for polyQ aggregation (Figure 7.2A). The unstructured polyQ region undergoes an unknown conformational conversion resulting in the formation of



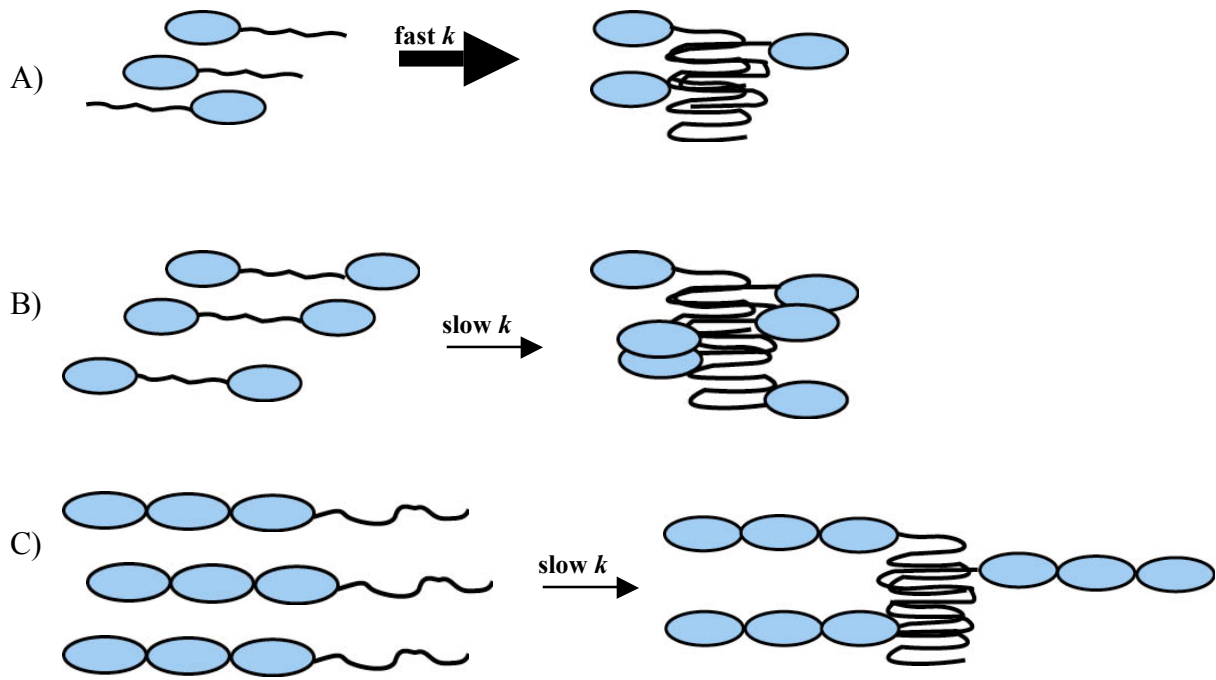
**Figure 7.1: Flanking domains can modulate the kinetic barrier to aggregation.** The native state (N) must overcome a kinetic barrier to convert into amyloid (A). The height of the kinetic barrier is influenced by properties of the flanking domain. *1.* Proteins that have interdomain polyQ tracts and/or multiple/large flanking domains have an increased kinetic barrier to aggregation and therefore a decreased rate of aggregation. *2.* Increasing the polyQ length lowers the kinetic barrier to aggregation. Also, the aggregation of flanking domains in polyQ proteins that have a multidomain aggregation mechanism lowers the kinetic barrier to polyQ aggregation. *3.* Flanking domain aggregation can be inhibited by sHsps which slow the rate of aggregation.

amyloid-like aggregates in which the polyQ region forms the core  $\beta$ -sheet of the fibril, decorated by folded domains (Figure 7.2A). Domains flanking fibril core regions can retain their folded and functional conformation in the fibrillar context (18, 36-38). The yeast prions Ure2p (36) and Sup35p (37), and a polyalanine model protein (38) retain their function upon misfolding and aggregation of adjacent amyloidogenic motifs. PolyQ repeats in the disease proteins are flanked by folded domains with functionality (143, 159, 213, 214). Unless proteolytic cleavage of a polyQ-containing fragment occurs prior to aggregation, the flanking domains will be involved in the fibril architecture and may retain their structured conformation and decorate the external

surface of the fibril core. Although there are no structural models of aggregates formed by full-length disease proteins, there is evidence indicating the maintenance of some normal cellular interactions of flanking domains in the fibrillar context (215).

Not all proteins with long polyQ tracts cause disease. For example, the non-pathological Forkhead box P2 contains a Q40 tract (216). The composition of regions flanking repeats can alter the biophysical behaviour of the polyQ region and adjust the kinetic barrier to amyloid formation (128, 129, 132). A recent crystallography-derived model of Htt exon1 showed that the composition of the flanking regions can induce some  $\alpha$ -helical and polyproline-II helical structure in the polyQ tract (212). As the biophysical properties of the polyQ tract are central to aggregation, and proteolytic cleavage is a trigger for aggregation, we hypothesized that in addition to local sequence elements, flanking domains can also modulate the biophysical properties of the polyQ region. If domains do remain attached and folded in the fibril context, what effect does this have on the polyQ aggregation kinetics? Our data showing that polyQ aggregation is suppressed by the presence of additional domains is the first experimental evidence to directly show a correlation between the number and location of domains and the rate of polyQ aggregation (Chapter 5).

Compared to the polyQ tract with a flexible, untethered C-terminus (Figure 7.2 A) we have shown that the rate of aggregation is significantly slowed when the polyQ tract is located between two identical domains (Figure 7.2 B). The lag phase is lengthened, suggesting that nucleation is less favourable, and a higher concentration of protein is required for aggregation to proceed, suggesting that the critical concentration required for aggregation is enhanced. Our data raise a number of questions: Does the domain architecture of the host protein influence the structure and/or dynamics of the polyQ region? Do adjacent domains sterically inhibit polyQ intermolecular interactions



**Figure 7.2: Proposed pathways of pathological polyQ aggregation.** The polyQ location and protein context affect the rate of aggregation. A) When a polyQ tract is flexible and untethered at one terminus, there is a fast rate ( $k$ ) of aggregation. B) When a polyQ is located between two domains aggregation occurs at a slow rate. C) The rate of aggregation is also slowed where there are multiple flanking domains.

and therefore elongation of aggregates? Kinetic modeling of nucleation and elongation would provide valuable insight into the mechanism of misfolding.

According to our database analysis, approximately 18 % of human non-disease polyQ-containing proteins have an interdomain repeat (Chapter 5). Repeat location may therefore provide a route by which polyQ-expanded variants of these proteins can avoid aggregation. Within the group of non-disease interdomain polyQ proteins, in most cases the repeat tract is flanked by two identical domains. Our experimental data show that there is a correlation between the number of repeating domains and the rate of aggregation, even when the domains are all N-terminal to the polyQ tract (Chapter 5, Figure 7.2 C). Together, these data suggest that genetic duplication of domains could be an evolutionary mechanism to avoid aggregation.

---

**7.13 Small Heat Shock Proteins and PolyQ Aggregation**

---

Aspects of the cellular environment can influence amyloid aggregation. Aggregation is modified by changes in solution conditions, macromolecular crowding, membrane interactions and protein-protein interactions (48, 170, 179, 217-219). Molecular chaperones are important cellular quality control proteins that assist proteins to reach and maintain their correctly folded conformation (220). The small heat shock proteins (sHsps) are a class of molecular chaperones that have a major role in preventing off-pathway protein aggregation and are therefore important in maintaining proteostasis (221).

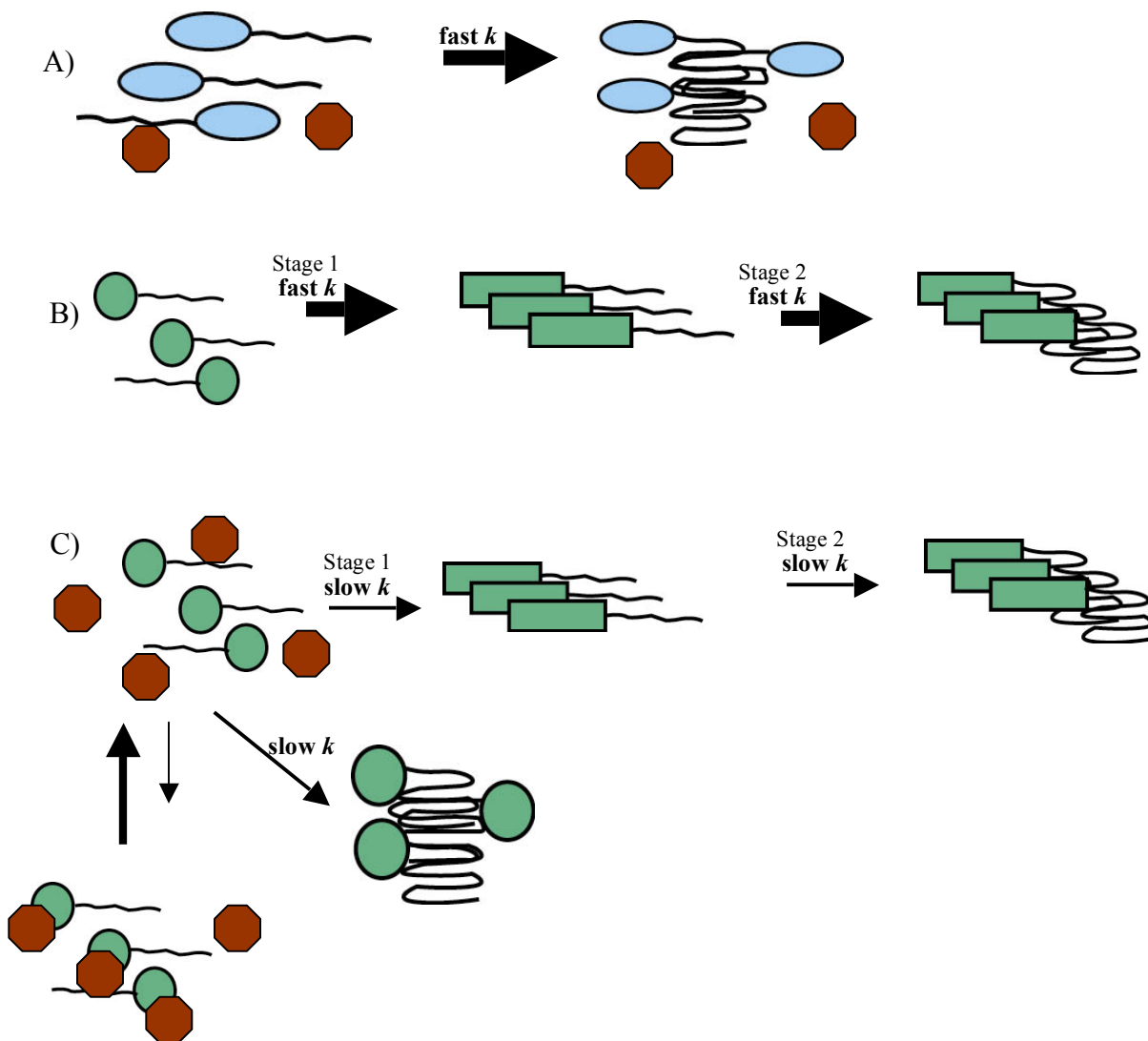
This thesis shows that the sHsp,  $\alpha$ B-crystallin, *does not* directly inhibit polyQ aggregation (Chapter 6, Fig 7.3 A). sHsps recognise misfolded substrates that have exposed hydrophobic regions (220, 221). Considering that glutamine side chains are intrinsically polar it is unlikely that sHsps recognise aggregation-prone polyQ sequences. In light of our data showing that a pathological length polyQ tract does not unfold the flanking domain in the context of the SpA-polyQ model protein (Chapter 4), the exposed hydrophobicity of the flanking domain would remain unchanged and the sHsp would not recognise the folded flanking domain.

A previous study from our laboratory has shown that self-association of the N-terminal Josephin domain is an important initial step in the *in vitro* aggregation pathway of ataxin-3 (Figure 7.3 B) (151). In this thesis we showed that  $\alpha$ B-crystallin *does* inhibit Josephin intermolecular interactions (Chapter 6). The polyQ aggregation stage is slowed, but not completely suppressed, and this is consistent with the observation that the sHsp does not directly prevent polyQ misfolding (Figure 7.3 C).

These data support the multidomain *in vitro* aggregation pathway for ataxin-3, and show that the rate of the polyQ-dependent aggregation stage can be modulated by



the ability of Josephin to initially self-associate. These data suggest that Josephin self-association in the context of ataxin-3 can lower the kinetic barrier to polyQ aggregation (Figure 7.1), possibly by increasing the local polyQ concentration. The presence of sHsps changes the energetic landscape such that Josephin aggregation becomes less favourable. Therefore sHsps can counteract the decreased kinetic barrier to polyQ aggregation caused by Josephin self-association (Figure 7.1).



**Figure 7.3. Pathways of polyQ aggregation and the role of sHsps.** A) In the presence of a sHsp polyQ aggregation proceeds at the same rate as in Fig 7.2 A. B) The multidomain aggregation pathway of ataxin-3 occurs by self-association of the N-terminal Josephin domain to form small soluble fibrils preceding aggregation of the polyQ regions. C) A sHsp inhibits the first stage of aggregation by transiently interacting with the aggregation-prone Josephin domain. This may induce an alternative aggregation pathway in which polyQ interactions occur in the absence of flanking domain interactions.

## **7.2 Native State Energetics and the Role of Flanking Domains**

Perturbations to the thermodynamic, kinetic or dynamic parameters of the native state modulate the propensity for aggregation. Mutations and the presence of amyloidogenic motifs can have a thermodynamic impact by disrupting the H-bonding network leading to partially-unfolded conformations. This thesis demonstrated that the H-bonding network can be affected by a polyQ tract when it is located within a domain (Chapter 4). If contacts between protein regions N- and C-terminal to the polyQ tract are necessary for overall stability, and these interactions are dependent on distance restraints that are overcome by a pathological length polyQ, H-bonds will be perturbed and an enthalpic penalty will result.

Considering the C-terminally fused polyQ tract did not affect the structure of the flanking domain, our data are consistent with theoretical models suggesting that polyQ aggregation is entropically-driven (125, 222). The lowest energy conformation of monomeric polyQ is random coil, which is stabilized by high configurational entropy (125). A highly unfavourable conformational conversion from random coil to a  $\beta$ -sheet rich monomer has been proposed to occur in the nucleation step that initiates aggregation (63). This reaction involves the formation of ordered structure, requiring a change in solvation state, in which the backbone solvates itself (rather than water as occurs in the random coil state) (125). Such a disorder to order transition requires the entropic barrier to be overcome. The ability to overcome this barrier is dependent on repeat length, as proteins containing a longer polyQ tract have a reduced aggregation lag time and hence faster rate of nucleation (Chapter 4).

What is the contribution of flanking regions to this entropic bottleneck? Theoretical models have shown that flexible, disordered peptide sequences fused to amyloidogenic motifs can perturb fibril formation by entropic interference, and the

extent of interference is dependent on the length of the flanking peptide (211). The effect of folded flanking domains on the overall chain thermodynamics has not been described, however, our data clearly show that the folded flanking domains in our polyQ system alter the kinetic barrier to fibril formation (Chapter 5). Larger proteins have slower translational diffusion and rotational relaxation rates compared to smaller proteins due to altered Brownian motion. In systems where proteins aggregate, altered molecular motion of larger proteins would affect the rate of conformational change and therefore alter nucleation, and would affect the frequency of collisions in solution, modulating the elongation rate (43). Consistent with this theory, multidomain SpA-polyQ proteins had slower nucleation and elongation rates compared with single domain variants (Chapter 5). The binding of molecular chaperones and other cellular proteins would affect the molecular size and by a similar theory this could contribute to altered aggregation rates. Also, additional aggregation stages, such as the Josephin self-association stage of ataxin-3 aggregation may enhance the rate of aggregation by restricting the conformational freedom and therefore reducing the entropy associated with native monomeric polyQ in addition to enhancing the local concentration of polyQ.

### **7.3 Further Questions about PolyQ Misfolding**

We have shown that a pathological length polyQ repeat has no effect on the stability or structure of an adjacent  $\alpha$ -helical flanking domain. As recombinant forms of the polyQ disease proteins are difficult to obtain, the structural properties of the domains flanking the polyQ tracts in most of the disease proteins have not been characterized. An ultimate goal would be to obtain recombinant variants of the disease proteins and determine their properties. The proteins may differ in stability and structural topology,

therefore it would be of value to determine the polyQ-induced misfolding mechanisms in variable protein contexts.

The results of this thesis suggest that protein domain architecture affects the biophysical behaviour of the polyQ tract and/or the rate of intersubunit interactions leading to fibril maturation. A recent crystallographic study of Htt exon1 has shown that the conformation of the polyQ tract is influenced by the immediately flanking regions (212). This has opened up the possibility of using high-resolution structural approaches to determine the behaviour of the polyQ tract in different protein contexts. Substituting the small domain used in our model protein with a larger domain may allow us to determine whether steric hindrance is a mechanism by which flanking domains can suppress polyQ aggregation.

If only the polyQ region comprises the fibril core, fibril stability would be largely determined by the enthalpic contribution of the H-bonding networks between the polyQ chains. This raises the question of whether there is a correlation between polyQ repeat length and fibril stability. Several studies have shown that proteins with pathological-length polyQ tracts form SDS-stable aggregates, however, it has not yet been shown whether stability increases with increasing repeat length (151, 223, 224). As only the polyQ region is involved in the fibril core of our model protein, this therefore provides an ideal system to determine fibril stability. A comparison of fibrils formed by the model protein and full-length ataxin-3 proteins would provide insight into the contribution of accessory domain aggregation to fibril stability.

Knowledge of the structural properties of endpoint aggregates and intermediate species is important, however, determining the initial native state perturbations and conformational events that trigger amyloid formation is critical in describing the molecular pathways of misfolding. As the initial trigger for polyQ aggregation is

thought to be a misfolded monomeric conformation, a major task would be to determine the polyQ conformational changes at the monomeric level. Ensemble experimental techniques are not capable of detecting such high-energy low-populous species. Furthermore, determining the structural properties of multimeric intermediates in the polyQ aggregation pathway, such as oligomers and protofibrils, will be critical in describing the disease mechanism as these intermediate species are proposed to be involved in toxicity. Resolving these questions will require ultrasensitive approaches such as single molecule fluorescence and NMR relaxation dispersion methods capable of detecting rarely populated species.

There are currently no therapies for the polyQ diseases. Ultimately the goal of research in this field is to identify therapeutic targets to avoid misfolding and aggregation. The results of this thesis suggest that targeting the conformational changes of the polyQ tract can prevent aggregation. The polyQ-binding peptide 1 (QBP1), and a number of small molecules, including the disaccharide trehalose (225), and the green tea polyphenol (-)-epigallocatechin-gallate (226) have been shown to decrease polyQ-derived aggregation and/or toxicity in cell and animal models. The next step would be to determine whether these compounds would be suitable human therapies. This thesis highlights the role of protein context in modulating aggregation. Non-polyQ protein regions therefore may provide potential therapeutic targets. Inhibitors that specifically interfere with properties of the non-polyQ domains such that; i) the native conformation is stabilized, ii) the solubility of flanking regions is enhanced, iii) proteolytic cleavage is reduced iv) aberrant protein interactions can be appropriately adjusted and/or v) the alternative steps involved in multidomain aggregation are inhibited may reduce the cellular toxicity and decrease the degenerative burden of the polyQ diseases.

## **Appendix 1**

**Polyglutamine fibril architecture is influenced by glutamine  
tract length and flanking regions**

---

## **Polyglutamine fibril architecture is influenced by glutamine tract length and flanking regions.**

Bronwen Thomas<sup>1</sup>, Amy L. Robertson<sup>1</sup>, Glyn Devlin<sup>1</sup>, Stephen P. Bottomley<sup>1</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Monash University, Victoria, Australia.

### **Abstract**

Polyglutamine (polyQ) diseases are associated with the formation of amyloid-like fibrils. These fibrils share many common structural properties, however the domains flanking the polyQ tract play an important role in aggregate formation. For example, in the case of ataxin-3 (atxn3), self-association of the globular Josephin domain precedes that of the polyQ tract. In this study we analysed the fibril structure of atxn3 variants containing polyQ tracts of pathogenic and non-pathogenic length and compared these to a model polyQ protein composed of non-aggregation-prone protein A fused to a polyQ tract (SpA-cQ). Utilizing Fourier transform infrared (FTIR) spectroscopy we demonstrate that the Josephin domain is structurally reorganised within the fibrils as compared to the soluble, monomeric proteins. This finding is in stark contrast to those acquired with the SpA-cQ proteins, which show protein A retains a native-like conformation within the fibrillar scaffold. Further to these findings, we show that fibrils composed of atxn3 bearing polyQ tract lengths in the pathological range are structurally distinct from those composed of non-pathogenic atxn3 variants. Cumulatively, these findings highlight the influence of both polyQ tract length and the properties of the flanking domains on polyQ protein aggregation.

**Key words:** Infrared spectroscopy, polyglutamine, ataxin-3

---

**Introduction**

Polyglutamines (polyQ) diseases are the most common form of inherited neurodegenerative diseases. The disease-related proteins show little homology beside the presence of a polyQ tract which is expanded above a threshold of approximately 40 repeats in most disease forms (Margolis and Ross 2001; Riley and Orr 2006). The hallmark of all the polyQ diseases is the presence of intranuclear inclusion bodies which have been identified in the nucleus, cytoplasm, dendrites, and axonal processes depending upon the specific disease (DiFiglia et al. 1997; Paulson et al. 1997; Schmidt et al. 1998; Fujigasaki et al. 2000). The importance of the polyQ tract in disease was eloquently demonstrated when transgenic mice carrying a non-disease related protein (hypoxanthine phosphoribosyltransferase) fused to an expanded polyQ tract developed symptoms characteristic of classic polyQ disease (Ordway et al. 1997).

PolyQ peptides are able to form fibres *in vitro*, that are comprised of pleated cross  $\beta$ -sheet (Perutz et al. 1993) and share many features typical of amyloid fibres (Chen et al. 2002). The cross  $\beta$ -sheet structure is characterised typically by the stacking together of two or more  $\beta$ -sheets along the fibril long axis, whereby the constituent  $\beta$ -strands are oriented perpendicular to this axis. One of the first structural models of polyQ aggregates suggested that they form polar zippers, whereby the cross  $\beta$ -sheet structure is stabilised not only by main-chain hydrogen bonding, but also by additional hydrogen bonds between by side-chain amides (Perutz et al. 1994). Recently, such polar zippers have been identified in fibrous crystals composed of short peptides with a range of primary sequences, although no such structure has been solved for polyQ peptides to date (Nelson et al. 2005; Nelson and Eisenberg 2006). Fibrils bearing this architecture give rise to  $\sim 4.7$  Å and  $\sim 10$  Å reflections in X-ray fibre diffraction patterns, arising from the inter-strand and inter-sheet spacings, respectively. The lack



of the  $\sim 10$  Å reflection in diffraction patterns of fibrils composed of short polyQ peptides prompted the modelling of an alternative polyQ fibril structure based on a  $\beta$ -helix, in which cylindrical  $\beta$ -sheets made of hydrogen bonded  $\beta$ -strands wound around the fibre axis (Perutz et al. 2002). This model has been disputed, however, with suggestions that the lack of  $\sim 10$  Å reflection may be due to lower total scattering intensity and greater background in the presence of water (Sikorski and Atkins 2005; Dzwolak et al. 2006; Shewmaker et al. 2006; Squires et al. 2006). Further models have been proposed, including parallel  $\beta$ -sheets (Shewmaker et al. 2006; Zanuy et al. 2006),  $\beta$ -hairpins (Sikorski and Atkins 2005), highly compact random coils (Starikov et al. 1999) and polyproline II helical structure (Chellgren et al. 2006), however, there currently remains a lack of experimental evidence to validate these models.

Short polyQ peptides provide a useful tool to study aggregation however, the naturally occurring polyQ proteins implicated in neurodegenerative diseases have their polyQ tract flanked by additional polypeptide sequence. On average the polyQ sequence comprises only  $\sim 1.5$ -20 % of the total molecular mass of the protein (Faux et al. 2005). It is unreasonable to expect that these flanking regions have no impact on the aggregation behaviour of the proteins and their resultant aggregate structures. Indeed, the interplay between folded flanking domains and the polyQ tract has been shown to play a central role in the aggregation propensity of the disease-related proteins ataxin-1 and ataxin-3 (atxn3) (Chow et al. 2004b; de Chiara et al. 2005). The pathway of atxn3 aggregation has recently been shown to be triggered by the aggregation-prone Josephin domain located N-terminal to the polyQ tract (Ellisdon et al. 2006; Thomas et al. In Preparation). Here, atxn3 aggregation is initiated by contacts made between Josephin domains, and only above a threshold polyQ length does the polyQ tract structurally consolidate on this initial aggregation event to give rise to more stable aggregates.

Similarly, recent findings have shown that the aggregation of huntingtin exon 1 is instigated by a 17 amino-acid region at the N-terminus, whose structure becomes increasingly perturbed by expansion of the neighbouring polyQ tract (Thakur et al. 2009).

The aim of this work was to examine the secondary structural properties of fibrils composed of atxn3 proteins with polyQ tract lengths in the pathogenic and non-pathogenic range by Fourier transform infrared (FTIR) spectroscopy. In addition, the fibril structure of a model protein, composed of the B domain of SpA attached to varying lengths of polyQ, was analysed. These data demonstrate that the model chimeric protein aggregates entirely via the polyQ tract with the SpA domain remaining structurally unchanged in the fibril. Further to this, the spectral properties of the  $\beta$ -sheets formed by the aggregated SpA-cQ variants were similar irrespective of polyQ tract length. This is in stark contrast to fibrils formed by atxn3 variants, which showed clear disruption of the Josephin domain in all cases and structural differences in the fibrils that were dependent on polyQ tract length.

---

**Results**

Five atxn3 variants and five SpA variants were analysed in total. Of the atxn3 variants studied, two contained polyQ tracts with length in the pathogenic range, atxn3(Q64) and atxn3(Q50), and a further two proteins contained tract lengths in the non-pathogenic range, atxn3(Q28) and atxn3(Q15) with 64, 50, 28 and 15 glutamine residues respectively. In addition to these the isolated N-terminal Josephin domain was also analysed. While, all atxn3 variants studied have been shown previously to aggregate, only the pathogenic variants form SDS-stable aggregates (Ellisdon et al. 2006). The SpA variants utilised were SpA-cQ0, spA-cQ35, spA-cQ40, spA-cQ48 and spA-cQ52 with 0, 35, 40, 48 and 52 glutamine residues respectively. All SpA-cQ variants with a polyQ tract longer than 35 aggregate to form SDS-stable fibrils (Robertson et al. 2008).

We first analysed the structure of monomeric and fibrillar SpA-cQ variants using FTIR spectroscopy (Figure 1). Absorbance by polypeptides in the amide I region ( $1600\text{-}1700\text{ cm}^{-1}$ ) arises predominantly from amide carbonyl stretching vibrations and as such is influenced by secondary structure. In the spectrum of monomeric SpA-cQ0 this region could be fitted to five Gaussian lineshapes, two of which corresponded to spectral bands arising from amino acid side-chains, with the remainder corresponding to  $\beta$ -turns, unordered and  $\alpha$ -helical secondary structural elements, with no  $\beta$ -sheet structure evident. The composition of secondary structure in SpA-cQ0 was determined by calculating the integral of the fitted band components, and yielded similar values to those obtained by deconvolution of far ultra violet circular dichroism (far-UV CD) spectra using the CONTINLL algorithm (Provencher and Glockner 1981; Van Stokkum et al. 1990; Robertson et al. 2008) (Supplementary material- Table 1). Together, these data are well-correlated with the predominantly  $\alpha$ -helical structure of SpA, as determined previously at high-resolution by NMR spectroscopy (Torigoe et al. 1990a;

Torigoe et al. 1990b). The high protein concentrations required for FTIR analysis in solution prevented the monomeric structures of the SpA variants with polyQ tracts from being assessed as aggregation occurred too rapidly. However, Robertson et al. (2008) demonstrated previously that the polyQ tract existed in a random coil conformation at lower concentrations.

Fibrils were produced from all of the SpA-cQ proteins, as described previously, for analysis by FTIR spectroscopy. The FTIR spectra of all SpA-cQ fibrils overlaid almost identically with each other (Figure 1B). The spectra were deconvoluted to determine the content of secondary structural elements in the fibrils. Each spectrum was fitted to six Gaussian lineshapes, five of which corresponded to secondary structural elements, with the additional band arising from amino acid side-chain vibrations (Figure 1C). The positions of the component bands in the fibril spectra were similar for all of the SpA-cQ variants, with no significant trend in the deviations (Table 1). Of the five bands corresponding to secondary structures, three were well-correlated with the structural elements observed for the monomeric protein (above), indicative that the SpA domain remains folded in a native-like conformation within the fibril, consistent with previous findings (Robertson et al. 2008). A major additional band in the fibril spectra was observed with a maximum at  $1626\text{ cm}^{-1}$  and a further minor band was present with a maximum at  $1675\text{ cm}^{-1}$ . The former component is commonly ascribed to the presence of  $\beta$ -strands (Susi et al. 1967; Susi and Byler 1987; Arrondo et al. 1993), and is consistent with the adoption of such a structure by the polyQ tract within the fibril. Indeed the  $1626\text{ cm}^{-1}$   $\beta$ -sheet peak location has previously been identified for polyQ peptides (Sharma et al. 1999). Furthermore, the position of this band at lower frequencies relative to those typically expected for  $\beta$ -sheet structures in globular proteins is indicative of fibrillar,  $\beta$ -sheet aggregates (Zandomeneghi et al.

2004; Cordeiro et al. 2006). The band at  $1675\text{ cm}^{-1}$  represents one of two bands in the fibril spectra typically assigned to  $\beta$ -turns, with the other located at  $\sim 1687\text{ cm}^{-1}$ . This is indicative of the presence of  $\beta$ -turns in at least two environments within the fibrils. The  $\beta$ -turn band at the higher wavenumber is similar to that observed in the monomeric protein. The band at lower wavenumbers, however, is unique to the fibril and has previously been associated with turns found within  $\beta$ -sheets (Arrondo et al. 1987; Muga et al. 1993; Arrondo et al. 1994). As such, this component likely represents turns between  $\beta$ -strands formed from the polyQ tract.

The specific content of secondary structural elements within the fibrils was determined and was found to vary with the length of the polyQ tracts in the different SpA-cQ variants (Figure 2D). Given that each of the SpA-cQ variants has a different polypeptide length, the fraction of secondary structure was converted to the number of amino acids per molecule per secondary structural element to allow for a meaningful comparison between variants. As detailed in figure 2D, there was no significant difference in  $\alpha$ -helix content in any of the fibrillar variants and the monomeric SpA-cQ0 ( $p < 0.01$ ), consistent with the maintenance of a native-like  $\alpha$ -helical fold of the SpA domain within the fibrils. Given that the SpA domain remains in a native-like fold within the fibril, and contains no  $\beta$ -sheet structure itself, this provides an ideal scenario for clearly evaluating the adoption of  $\beta$ -sheet structure by the poly-Q tract. In the case of SpA-cQ35, which contains 35 glutamine residues, only 23 residues were observed to adopt  $\beta$ -sheet structure within the fibril, with the remaining glutamine residues seemingly adopting  $\beta$ -turn structure. Whilst a strong correlation between increasing  $\beta$ -sheet content and increasing polyQ tract length was observed ( $R = 0.96$ ,  $p < 0.05$ ), only four additional glutamine residues were incorporated into  $\beta$ -sheet in the SpA-cQ52 variant bearing 52 glutamine residues. These findings highlight a redundancy in the  $\beta$ -

sheet content relative to the increasing polyQ tract length, and demonstrates in this case, that a large proportion of the glutamine residues in the tract adopt  $\beta$ -turn and unordered structures, rather than  $\beta$ -strands. Further to these findings, for the shorter SpA-cQ variants in the non-pathological polyQ tract range (SpA-cQ35 and 40), a small number of residues that adopt unordered structure in the natively folded SpA domain conformationally convert to adopt  $\beta$ -turn and/or  $\beta$ -sheet structure in the fibrils. We infer from these findings that the polyQ tracts with lengths below the disease threshold may not bear sufficient conformational freedom to adopt a stable  $\beta$ -sheet structure within the fibril, and that additional residues from the C-terminus of the SpA domain are utilised to lend this additional freedom to the polyQ-tract.

We next turned our focus to the atxn3 proteins. Analysis of the secondary structure of native atxn3 proteins by deconvolution of the FTIR spectra (Figure 2) revealed similar results to those determined by deconvolution of previously published far-UV CD spectra (see supplementary material- Table 2) (Chow et al. 2004a; Masino et al. 2004; Ellisdon et al. 2006; Thomas et al. In Preparation). The Josephin domain of atxn3 was shown to contain predominantly  $\alpha$ -helical and  $\beta$ -sheet/turn structure, consistent with the high-resolution NMR structure of this domain (Mao et al. 2005; Nicastro et al. 2005). Unsurprisingly, the atxn3 variants contained significantly more unordered structure than the Josephin domain, owing to the presence of the polyQ tracts and additional sequence elements that are predicted to be unordered (Masino et al. 2003). The FTIR spectra of the atxn3 fibrils were all fitted by a minimum of six Gaussian lineshapes, however the fraction of the structure attributable to  $\alpha$ -helices and unordered structures could not be distinguished within the fibrils due to the spectral overlap of these bands and the structural complexity of the species (Surewicz et al. 1993; Arrondo et al. 1994; Lin and Chu 2003) (Figure 3).

The atxn3 fibril spectra, as compared to those acquired for the monomeric proteins, contained an additional  $\beta$ -strand band at  $\sim 1620\text{ cm}^{-1}$ , indicative of the transition to aggregated  $\beta$ -sheet structures. Further to this, the other band positions within the fibril spectra that were attributable to secondary structural elements were significantly shifted relative to those observed in the spectra of atxn3 monomers. This finding is in contrast to results from the SpA-cQ system, in which the SpA domain remains structurally unchanged with the fibril, and strongly indicates that the majority of the atxn3 polypeptide, and not simply the polyQ tract, undergoes substantial conformational reorganisation upon incorporation into the fibril. This is consistent with previous studies showing that the folded Josephin domain of atxn3 is at least partially disrupted within the aggregate, and is to some extent incorporated into the  $\beta$ -sheet core of the fibrils (Ellisdon et al. 2006).

In contrast to the findings for the SpA-cQ fibrils, the band positions corresponding to secondary structural elements in the FTIR spectra of the atxn3 variants shifted as a function of polyQ tract length (Figure 3B), indicating that the fibril structure changes subtly as the tract expands. There was a significant shift,  $\sim 7\text{ cm}^{-1}$ , to higher wavenumbers in the bands at  $\sim 1670$  and  $1685\text{ cm}^{-1}$ , corresponding to  $\beta$ -turns, as the polyQ tract length increased ( $p < 0.05$ ) (Table 2; Supplementary Table 2). A similar trend, albeit of a smaller magnitude, was observed with respect to the location of the aggregated  $\beta$ -sheet band ( $\sim 1620\text{ cm}^{-1}$ ). This band reproducibly shifted in all experiments by  $\sim 1.5\text{ cm}^{-1}$  to higher wavenumbers as the polyQ tract increased in length. The shift in the bands corresponding to aggregated  $\beta$ -sheets and  $\beta$ -turns are maximal as the length of the polyQ tract increases from 15 to 50 residues, with the spectral changes remaining relatively stable between 50 to 64 residues. Indeed, the shifts in these bands observed between atxn3(Q28) and atxn3(Q50) are statistically significant ( $p < 0.05$ ),

however the shifts observed between atxn3(Q50) and atxn3(Q64) are not ( $p < 0.05$ ). These data imply that fibril structure is more sensitive to polyQ tract length in the non-pathological range, and that once the polyQ tract reaches a threshold length, a specific fibril structure is maintained.

The number of amino acids, per molecule, contributing to each secondary structural element in the atxn3 fibrils was calculated from the integrals of the bands fitted to the FTIR spectra (Fig. 3C). Not surprisingly, significant differences can be seen between Josephin and full length atxn3 with respect to all secondary structure elements, given that Josephin is composed of only 182 amino acids, while the full length proteins are composed of 374-423 amino acids. The number of amino acids contributing to both  $\beta$ -sheets and  $\alpha$ -helices/unordered structures in the fibrils significantly ( $p < 0.01$ ) increased as the polyQ tract length expanded, whilst the number of amino acids contributing to  $\beta$ -turns decreased. There was no significant ( $p < 0.01$ ) difference in the number of amino acids contributing to  $\beta$ -sheets in the atxn3(Q64) and atxn3(Q50) fibrils, nor was there a significant difference between atxn3(Q15) and atxn3(Q28) variants. There was, however, a significant difference with respect to the number amino acids contributing to  $\beta$ -sheets between these two groups. Taken together with our findings that there are distinct spectral shifts in the component bands between fibrils composed of atxn3 variants with shorter and expanded polyQ tracts, these data further indicate that there are subtle structural differences in the fibrils produced from non-pathogenic and pathogenic length atxn3 variants.



---

**Discussion**

Presently there are over 40 defined protein conformational diseases arising from specific peptides or proteins failing to fold or remain folded (Chiti and Dobson 2006). This ‘misfolding’ leads to aggregate formation, often resulting in significant cellular dysfunction (Stefani and Dobson 2003). The amyloidogenic proteins implicated in these diseases differ widely with respect to primary sequence and native fold, however the fibrillar structures they assemble into show great similarities due to a common core structure consisting of highly ordered  $\beta$ -sheets (Rochet and Lansbury Jr 2000; Knowles et al. 2007). A subset of these amyloidogenic proteins contain polyQ tracts and are associated with nine inherited diseases including Huntington disease, spinal and bulbar muscular atrophy (SBMA) and Machado Joseph disease (MJD).

Given the prevalence of these diseases, a great deal of research effort has gone into investigating the mechanisms of polyQ protein aggregation. Many of these investigations have focussed solely on short polyQ peptides as the polyQ tract is considered to confer aggregation propensity to the polyQ proteins. Of the disease related polyQ proteins, however, the polyQ tracts comprise only a small fraction of their total molecular mass (Faux et al. 2005), and non-glutamine sequence elements have recently been shown to be capable of triggering the aggregation of these proteins (de Chiara et al. 2005; Ellisdon et al. 2006; Thakur et al. 2009). In the case of atxn3 the Josephin domain N-terminal to the polyQ tract forms initial contacts within the aggregates, which are subsequently consolidated by more stable polyQ mediated contacts (Ellisdon et al. 2006; Thomas et al. In Preparation). It is anticipated that such involvement by regions flanking the polyQ tract in the aggregation mechanism of this protein will influence the structures of the aggregates, and as such we examined the structure of these fibrils using FTIR. As a contrast to these studies, we also investigated

the structures of fibrils composed of SpA-cQ proteins, as previous data has indicated that the SpA domain at the N-terminus of these proteins does not play a role in their aggregation mechanism.

The FTIR data presented here strongly indicates that the SpA domain retains a native-like conformation upon incorporation of the protein into fibrils given that the number of amino acids incorporated into  $\alpha$ -helices does not significantly change for any of the variants compared to the monomer. As fibrils composed of simple polyQ peptides have previously been shown to contain predominantly  $\beta$ -sheet structure with no  $\alpha$ -helical content (Perutz et al. 1994; Sharma et al. 1999), we conclude from these data that the SpA domain is not perturbed during aggregation and is not incorporated into the  $\beta$ -sheet core of the fibrils. Indeed further support for this finding is the location of the  $\beta$ -sheet peak within the FTIR spectra of the SpA-cQ fibrils at  $\sim 1626\text{ cm}^{-1}$ , which is almost identical to that previously identified for polyQ peptides (Sharma et al. 1999). Together these findings indicate not only that the SpA domain is not involved in the fibril, but that its presence does not disturb the structure of the  $\beta$ -sheets formed by the polyQ tracts.

Comparatively, the FTIR data indicated that the structure of the Josephin domain was perturbed in fibrils composed of the atxn3 variants relative to the soluble, monomeric proteins. Although the structure of the Josephin domain within the atxn3 fibrils could not be completely resolved owing the structural complexity of the domain, the shifts of the component bands in the FTIR spectra upon conversion of the soluble protein into fibrils indicates substantial structural reorganisation of the entire protein. Further to this, the  $\beta$ -sheet spectral band arising from the atxn3 fibrils was located between  $1617.5\text{-}1621.1\text{ cm}^{-1}$  depending on the length of the polyQ tract. This is significantly red-shifted relative to the  $\beta$ -sheet band previously observed for fibrils

composed of polyQ homopolymers ( $1627\text{ cm}^{-1}$ ) (Sharma et al. 1999), indicative of the incorporation of non-polyQ sequence elements into the  $\beta$ -sheet core of these fibrils.

The aggregation of atxn3, irrespective of polyQ tract length, is initiated by interactions formed between Josephin domains, giving rise to ThioT-reactive, SDS-soluble aggregates (Thomas et al. In Preparation). Pathogenic length atxn3 is able to undergo an additional step to form SDS-stable aggregates which are morphologically distinct from the earlier curvi-linear structures. (Ellisdon et al. 2006). Fibrils composed of atxn3 with polyQ tracts in the pathogenic and non-pathogenic length range appear to be structurally distinct by FTIR, with significant differences in the positions of the  $\beta$ -sheet bands and their total  $\beta$ -sheet content. The increased number of residues, per molecule, incorporated into the fibrillar,  $\beta$ -sheet core of the pathogenic variants is consistent with observations that these variants give rise to more stable aggregates. Similarly, the SpA variants with polyQ tracts in the pathogenic range contained higher  $\beta$ -sheet content than those with shorter polyQ tracts. However, in contrast to the atxn3 fibrils, the  $\beta$ -sheet band position of the spectra of SpA-cQ fibrils remained stable at  $\sim 1626\text{ cm}^{-1}$  for all variants, indicating that these fibrils shared a very similar overall architecture, regardless of polyQ tract length. Taken altogether, these data suggest that atxn3 fibril architecture is dependent on a subtle interplay between the Josephin domain and the polyQ tract length.

The two fibril systems demonstrate the fibrillar differences resulting from polyQ aggregation in the presence and absence of an aggregation-prone domain. While the stabilities of SpA and Josephin are similar, the former is stable in solution and has significantly reduced aggregation propensity (Chow et al. 2004c; Robertson et al. 2008). This result highlights the important role the flanking sequences play in directing polyQ mediated aggregation and sheds light on the varied results obtained by differing model

---

proteins (Stott et al. 1995; Tanaka et al. 2001; Bulone et al. 2006; Ignatova and Gierasch 2006; Robertson et al. 2008).

This study demonstrates the importance of the aggregation propensity of flanking domains when analysing polyQ aggregation and particularly when designing model chimeras to emulate disease-related proteins. Importantly, these data support previous work that the stability of the SpA domain is not disrupted by the addition of expanded polyQ tracts (Robertson et al. 2008) and thus does not get incorporated into the subsequent fibrils formed. In addition, this work supported the role of Josephin during atxn3 aggregation (Ellisdon et al. 2006; Thomas et al. In Preparation) and proposes that in the presence of expanded polyQ tracts the initial interactions dictated by Josephin are subsequently altered to allow for the formation of more stable polyQ mediated aggregates.

---

**Materials and Methods****Expression and Purification of Proteins**

All atxn3 variants were expressed as described previously (Chow et al. 2004a). The purification technique was similar to that published previously with minor improvements including the final buffer containing no glycerol (Thomas et al. In Preparation). The SpA variants were expressed and purified as previously described (Robertson et al. 2008). Protein was stored at -80 °C and run on a Superose 12 column prior to each experiment to confirm it was monomeric.

**Circular Dichroism**

The atxn3 proteins were added at 1 mg/ml to a buffered solution containing 100 mM Tris (pH 7.4), 80 mM NaCl solution. The SpA-cQ variants were added at 1mg/ml to a buffering solution containing 20 mM sodium acetate (pH 5.0), 100 mM NaCl, and 10 % (v/v) glycerol. The solutions were briefly vortexed and left to equilibrate for 5 minutes. All far-UV CD spectra were measured on a Jasco-810 spectropolarimeter with a path length of 0.1 mm. Spectra were recorded from 190 to 260 nm, using a 5-s/point signal averaging, and a scan speed of 50 nm/min. Spectra were analysed using the continLL algorithm, DICHROWEB (Whitmore and Wallace 2004; Whitmore and Wallace 2008)

**Fourier Transform Infrared Spectroscopy**

Monomeric proteins were analysed in the buffer specified for far-UV CD while protein aggregates were centrifuged and washed with ddH<sub>2</sub>O repeatedly prior to analysis. Spectra were recorded on a Bruker Bio-Attenuated Total Reflectance (ATR) cell II using a Bruker IFS 55 Equinox Fourier transform infrared (FTIR) spectrometer equipped with a liquid nitrogen cooled mercury cadmium telluride detector and a single

---

bounce silicon internal reflection element (IRE). For each spectrum at least 1 mg of protein was deposited onto the IRE and 256 interferograms were co-added at 2 cm<sup>-1</sup> resolution. Either a water background, or buffer background was subtracted from the protein spectra depending on the nature of the sample. The second derivatives of the spectra were taken using the operating software (OPUS NT version 5.0) in order to identify the positions of component bands. Raw spectra were subsequently deconvoluted by fitting Gaussian lineshapes to the component bands using the Origin<sup>®</sup> 7 software package (OriginLab Corporation, MA, USA). Spectra of at least three samples of each protein were acquired in independent experiments. Each spectrum was fitted independently and the results of these fits averaged.

## References

- Arrondo, J.L., Mantsch, H.H., Mullner, N., Pikula, S., and Martonosi, A. 1987. Infrared spectroscopic characterization of the structural changes connected with the E1---E2 transition in the Ca<sup>2+</sup>-ATPase of sarcoplasmic reticulum. *J Biol Chem.* **262**: 9037-9043.
- Arrondo, J.L.R., Castresana, J., Valpuesta, J.M., and Goni, F.M. 1994. Structure and thermal denaturation of crystalline and noncrystalline cytochrome oxidase as studied by infrared spectroscopy. *Biochemistry* **33**: 11650 - 11655.
- Arrondo, J.L.R., Muga, A., Castresana, J., and Goñi, F.M. 1993. Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy *Prog. Biophys. Mol. Biol.* **59**: 23.
- Bulone, D., Masino, L., Thomas, D.J., San Biagio, P.L., and Pastore, A. 2006. The interplay between polyQ and protein context delays aggregation by forming a reservoir of protofibrils. *PLoS ONE* **1**: e111.
- Chellgren, B.W., Miller, A.-F., and Creamer, T.P. 2006. Evidence for polyproline II helical structure in short polyglutamine tracts. *J. Mol. Biol.* **361**: 362-371.
- Chen, S., Berthelie, 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.
- Chiti, F., and Dobson, C. 2006. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**: 333-366.
- Chow, M.K., Ellisdon, A.M., Cabrita, L.D., and Bottomley, S.P. 2004a. Polyglutamine expansion in ataxin3 does not affect protein stability. *J. Biol. Chem.* **279**: 47643-47651.
- Chow, M.K., Paulson, H.L., and Bottomley, S.P. 2004b. Destabilization of a non-pathological variant of ataxin-3 results in fibrillogenesis via a partially folded intermediate: a model for misfolding in polyglutamine disease. *J. Mol. Biol.* **335**: 333-341.
- Chow, M.K.M., Mackay, J.P., Whisstock, J.C., Scanlon, M.J., and Bottomley, S.P. 2004c. Structural and functional analysis of the Josephin domain of the polyglutamine protein ataxin-3. *Biochem. Biophys. Res. Commun.* **322**: 387-394.
- Cordeiro, Y., Kraineva, J., Suarez, M.C., Tempesta, A.G., Kelly, J.W., Silva, J.L., Winter, R., and Foguel, D. 2006. Fourier transform infrared spectroscopy provides a fingerprint for the tetramer and for the aggregates of transthyretin. *Biophys. J.* **91**: 957-967.
- de Chiara, C., Menon, R.P., Dal Piaz, F., Calder, L., and Pastore, A. 2005. Polyglutamine is not all: The functional role of the AXH domain in the ataxin-1 protein. *J. Mol. Biol.* **354**: 883-893.
- DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., and Aronin, N. 1997. Aggregation of Huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**: 1990-1993.
- Dzwolak, W., Lokszejn, A., and Smirnovas, V. 2006. New insights into the self-assembly of insulin amyloid fibrils: An H-D exchange FT-IR study. *Biochemistry* **45**: 8143-8151.
- Ellisdon, A.M., Thomas, B., and Bottomley, S.P. 2006. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J. Biol. Chem.* **281**: 16888-16896.
- Faux, N.G., Bottomley, S.P., Lesk, A.M., Irving, J.A., Morrison, J.R., de la Banda, M.G., and Whisstock, J.C. 2005. Functional insights from the distribution and role of homeopeptide repeat-containing proteins. *Genome Res* **15**: 537-551.

- Fujigasaki, H., Uchihara, T., Koyano, S., Iwabuchi, K., Yagishita, S., Makifuchi, T., Nakamura, A., Ishida, K., Toru, S., Hirai, S., et al. 2000. Ataxin-3 is translocated into the nucleus for the formation of intranuclear inclusions in normal and Machado-Joseph disease brains. *Exp. Neurol.* **165**: 248-256.
- Ignatova, Z., and Gierasch, L.M. 2006. Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein. *J. Biol. Chem.* **281**: 12959-12967.
- Knowles, T.P., Fitzpatrick, A.W., Meehan, S., Mott, H.R., Vendruscolo, M., Dobson, C.M., and Welland, M.E. 2007. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science* **318**: 1900-1903.
- Lin, S.-Y., and Chu, H.-L. 2003. Fourier transform infrared spectroscopy used to evidence the prevention of [beta]-sheet formation of amyloid [beta](1-40) peptide by a short amyloid fragment. *Int. J. Biol. Macromol.* **32**: 173-177.
- Mao, Y., Senic-Matugilia, F., Di Fiore, P.P., Polo, S., Hodsdon, M.E., and De Camilli, P. 2005. Deubiquitinating function of ataxin-3: Insights from the solution structure of Josephin domain. *Proc. Natl. Acad. Sci. USA* **102**: 12700-12705.
- Margolis, R.L., and Ross, C.A. 2001. Expansion explosion: new clues to the pathogenesis of repeat expansion neurodegenerative diseases. *Trends Mol. Med.* **7**: 479-482.
- Masino, L., Musi, V., Menon, R.P., Fusi, P., Kelly, G., Frenkiel, T.A., Trottier, Y., and Pastore, A. 2003. Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS Lett.* **549**: 21-25.
- Masino, L., Nicastro, G., Menon, R.P., Piaz, F.D., Calder, L., and Pastore, A. 2004. Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3. *J Mol Biol* **344**: 1021-1035.
- Muga, A., Arrondo, J.L.R., Bellon, T., Sancho, J., and Bernabeu, C. 1993. Structural and functional studies on the interaction of sodium dodecyl sulfate with  $\beta$ -Galactosidase. *Arch. Biochem. Biophys.* **300**: 451-457.
- Nelson, R., and Eisenberg, D. 2006. Recent atomic models of amyloid fibril structure. *Curr. Opin. Struct. Biol.* **16**: 260-265.
- Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riek, C., Grothe, R., and Eisenberg, D. 2005. Structure of the cross- $\beta$  spine of amyloid-like fibrils. *Nature* **435**: 773-778.
- Nicastro, G., Menon, R.P., Masino, L., Knowles, P.P., McDonald, N.Q., and Pastore, A. 2005. The solution structure of the Josephin domain of ataxin-3: Structural determinants for molecular recognition. *Proc. Natl. Acad. Sci. USA* **102**: 10493-10498.
- Ordway, J.M., Tallaksen-Greene, S., Gutekunst, C.A., Bernstein, E.M., Cearley, J.A., Wiener, H.W., Dure, L.S.t., Lindsey, R., Hersch, S.M., Johe, R.S., et al. 1997. Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell* **91**: 753-763.
- 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.
- Perutz, M., Johnson, T., Suzuki, M., and Finch, J. 1994. Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. USA* **91**: 5355-5358.



- Perutz, M.F., Finch, J.T., Berriman, J., and Lesk, A. 2002. Amyloid fibers are water-filled nanotubes. *Proc Natl Acad Sci USA* **99**: 5591-5595.
- Perutz, M.F., Staden, R., Moens, L., and De Baere, I. 1993. Polar zippers. *Curr. Biol.* **3**: 249-253.
- Provencher, S.W., and Glockner, J. 1981. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**: 33-37.
- Riley, B.E., and Orr, H.T. 2006. Polyglutamine neurodegenerative diseases and regulation of transcription: assembling the puzzle. *Genes Dev.* **20**: 2183-2192.
- Robertson, A.L., Horne, J., Ellisdon, A.M., Thomas, B., Scanlon, M.J., and Bottomley, S.P. 2008. The structural impact of a polyglutamine tract is location- dependent. *Biophys. J.* **95**: 5922-5930.
- Rochet, J.-C., and Lansbury Jr, P.T. 2000. Amyloid fibrillogenesis: themes and variations. *Curr. Opin. Struct. Biol.* **10**: 60-68.
- Schmidt, T., Landwehrmeyer, G.B., Schmitt, I., Trotter, Y., Auburger, G., Laccone, F., Klockgether, T., Volpel, M., Epplen, J.T., Schols, L., et al. 1998. An isoform of ataxin-3 accumulates in the nucleus of neuronal cells in affected brain regions of SCA3 patients. *Brain Pathol* **8**: 669-679.
- Sharma, D., Sharma, S., Pasha, S., and Brahmachari, S.K. 1999. Peptide models for inherited neurodegenerative disorders: conformation and aggregation properties of long polyglutamine peptides with and without interruptions. *FEBS Lett.* **456**: 181-185.
- Shewmaker, F., Wickner, R.B., and Tycko, R. 2006. Amyloid of the prion domain of Sup35p has an in-register parallel beta-sheet structure. *Proc. Natl. Acad. Sci. USA* **103**: 19754-19759.
- Sikorski, P., and Atkins, E. 2005. New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils. *Biomacromol.* **6**: 425-432.
- Squires, A.M., Devlin, G.L., Gras, S.L., Tickler, A.K., MacPhee, C.E., and Dobson, C.M. 2006. X-ray scattering study of the effect of hydration on the cross-beta structure of amyloid fibrils. *J Am Chem Soc* **128**: 11738-11739.
- Starikov, E.B., Lehrach, H., and Wanker, E.E. 1999. Folding of oligoglutamines: a theoretical approach based upon thermodynamics and molecular mechanics. *J. Biomol. Struct. Dyn.* **17**: 409-427.
- Stefani, M., and Dobson, C.M. 2003. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *Journal of Molecular Medicine* **81**: 678-699.
- 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. USA* **92**: 6509-6513.
- Surewicz, W.K., Mantsch, H.H., and Chapman, D. 1993. Determination of protein secondary structure by Fourier transform infrared spectroscopy: A critical assessment. *Biochemistry* **32**: 389-394.
- Susi, H., and Byler, D.M. 1987. Fourier transform infrared study of proteins with parallel beta-chains. *Arch Biochem Biophys* **258**: 465-469.
- Susi, H., Timasheff, S.N., and Stevens, L. 1967. Infrared spectra and protein conformations in aqueous solutions. The amide I band in H<sub>2</sub>O and D<sub>2</sub>O solutions. *J. Biol. Chem.* **242**: 5460-5466.
- Tanaka, M., Morishima, I., Akagi, T., Hashikawa, T., and Nukina, N. 2001. Intra- and intermolecular  $\beta$ -pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases. *J. Biol. Chem.* **276**: 45470-45475.

- Thakur, A.K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V.M., Byeon, I.-J.L., Anjum, D.H., Kodali, R., Creamer, T.P., Conway, J.F., et al. 2009. Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat. Struct. Mol. Biol.* **Advanced Publication**.
- Thomas, B., Kerr, M.L., Ellisdon, A.M., and Bottomley, S.P. In Preparation. Identification of pre-fibrillar intermediates in ataxin-3 fibrillogenesis.
- Torigoe, H., Shimada, I., Saito, A., Sato, M., and Arata, Y. 1990a. Sequential <sup>1</sup>H NMR assignments and secondary structure of the B domain of staphylococcal protein A: structural changes between the free B domain in solution and the Fc-bound B domain in crystal. *Biochemistry* **29**: 8787-8793.
- Torigoe, H., Shimada, I., Waelchli, M., Saito, A., Sato, M., and Arata, Y. 1990b. <sup>15</sup>N nuclear magnetic resonance studies of the B domain of staphylococcal protein A: sequence specific assignments of the imide <sup>15</sup>N resonances of the proline residues and the interaction with human immunoglobulin G. *FEBS Lett* **269**: 174-176.
- Van Stokkum, I.H.M., Spoelder, H.J.W., Bloemendal, M., Van Grondelle, R., and Groen, F.C.A. 1990. Estimation of protein secondary structure and error analysis from CD spectra. *Anal. Biochem* **191**: 110-118.
- Whitmore, L., and Wallace, B.A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**: 668-673
- Whitmore, L., and Wallace, B.A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **89**: 392-400.
- Zandomenoghi, G., Krebs, M.R.H., McCammon, M.G., and Fandrich, M. 2004. FTIR reveals structural differences between native  $\beta$ -sheet proteins and amyloid fibrils. *Protein Sci.* **13**: 3314-3321.
- Zanuy, D., Gunasekaran, K., Lesk, A.M., and Nussinov, R. 2006. Computational study of the fibril organization of polyglutamine repeats reveals a common motif identified in  $\beta$ -helices. *J. Mol. Biol.* **358**: 330-345.

**Table 1. Peak locations for Protein A fibrils**

The values are shown as mean (ave.) and standard deviation (S.D.) obtained from at least four separate experiments.

|                                    | pAQ35  |      | pAQ40  |      | pAQ48  |      | pAQ52  |      |
|------------------------------------|--------|------|--------|------|--------|------|--------|------|
|                                    | Ave.   | S.D. | Ave.   | S.D. | Ave.   | S.D. | Ave.   | S.D. |
| <b><math>\beta</math>-Sheets</b>   | 1626.0 | 0.7  | 1626.7 | 0.1  | 1626.1 | 0.7  | 1626.3 | 0.1  |
| <b>Unordered</b>                   | 1641.2 | 0.5  | 1642.5 | 0.2  | 1642.2 | 0.5  | 1642.7 | 0.2  |
| <b><math>\alpha</math>-Helices</b> | 1657.8 | 0.2  | 1658.4 | 0.1  | 1657.9 | 0.2  | 1657.9 | 0.3  |
| <b><math>\beta</math>-Turns</b>    | 1673.5 | 0.3  | 1674.4 | 0.2  | 1673.5 | 0.3  | 1672.7 | 0.0  |
| <b><math>\beta</math>-Turns</b>    | 1687.5 | 0.2  | 1688.6 | 0.5  | 1688.1 | 0.2  | 1687.3 | 0.4  |

**Table 2. Peak locations for ataxin-3 fibrils**

The values are shown as mean (ave.) and standard deviation (S.D.) obtained from at least four separate experiments.

|                                 | Josephin |     | Atxn3(Q15) |     | Atxn3(Q28) |     | Atxn3(50) |     | Atxn3(Q64) |     |
|---------------------------------|----------|-----|------------|-----|------------|-----|-----------|-----|------------|-----|
|                                 | Ave.     | S.D | Ave.       | S.D | Ave.       | S.D | Ave.      | S.D | Ave.       | S.D |
| <b>β-Sheets</b>                 | 1617.5   | 0.8 | 1619.7     | 0.3 | 1619.7     | 0.4 | 1620.9    | 0.3 | 1621.1     | 0.3 |
| <b>β-Sheets</b>                 | 1637.9   | 1.0 | 1636.9     | 0.5 | 1637.8     | 0.4 | 1637.1    | 0.3 | 1638.4     | 0.6 |
| <b>α-Helices/<br/>Unordered</b> | 1657.1   | 1.4 | 1653.1     | 1.5 | 1655.0     | 0.8 | 1656.8    | 0.4 | 1657.8     | 0.3 |
| <b>β-Turns</b>                  | 1673.0   | 1.3 | 1667.8     | 1.9 | 1670.8     | 2.2 | 1674.4    | 0.7 | 1675.4     | 0.8 |
| <b>β-Turns</b>                  | 1686.2   | 1.4 | 1681.3     | 0.0 | 1684.7     | 2.6 | 1687.3    | 1.1 | 1687.5     | 0.3 |

---

**Figure Legends****Figure 1. Secondary structure of monomeric and aggregated protein A**

FTIR spectra of monomeric SpA-cQ0 were composed of  $\beta$ -turns (1),  $\alpha$ -helices (2) and unordered structures (3) (A). FTIR spectra of aggregated SpA-cQ35 (light grey), SpA-cQ40 (grey), SpA-cQ48 (dashed black) and SpA-cQ52 (black) (B). The spectra of the aggregates were all composed of  $\beta$ -turns (1),  $\beta$ -turns (2),  $\alpha$ -helices (3), unordered structures (4) and  $\beta$ -sheets (5) and a fitted example of SpA-cQ35 is shown (C). The spectral information below  $1620\text{ cm}^{-1}$  is obtained from side-chain interactions and the amide II region. The average number of amino acids, per molecule, involved in each secondary structure within the aggregates was determined and the values are shown with standard error (D).

**Figure 2. Secondary structure ataxin-3 monomers**

The FTIR spectra of monomeric atxn3(Q64) (black), atxn3(Q50) (dark grey), atxn3(Q28) (grey), atxn3(Q15) (light grey) and Josephin (black dashed) were obtained (A). The spectra were all fitted to six peaks composed of  $\beta$ -turns (1),  $\beta$ -turns (2),  $\alpha$ -helices (3), unordered (4)  $\beta$ -sheets (5),  $\beta$ -sheets (5) and side chains (B).

**Figure 3. Secondary structure of ataxin-3 aggregates**

The FTIR spectra of atxn3(Q64) (black), atxn3(Q50) (dark grey), atxn3(Q28) (grey), atxn3(Q15) (light grey) and Josephin (black dashed) were obtained (A). The spectra were all fitted to six peaks composed of  $\beta$ -turns (1),  $\beta$ -turns (2),  $\alpha$ -helices/unordered (3)  $\beta$ -sheets (4),  $\beta$ -sheets (5) and side-chains (B). The number of amino acids, per molecule, involved in each secondary structure was determined from the peak areas (C).

Figure 1

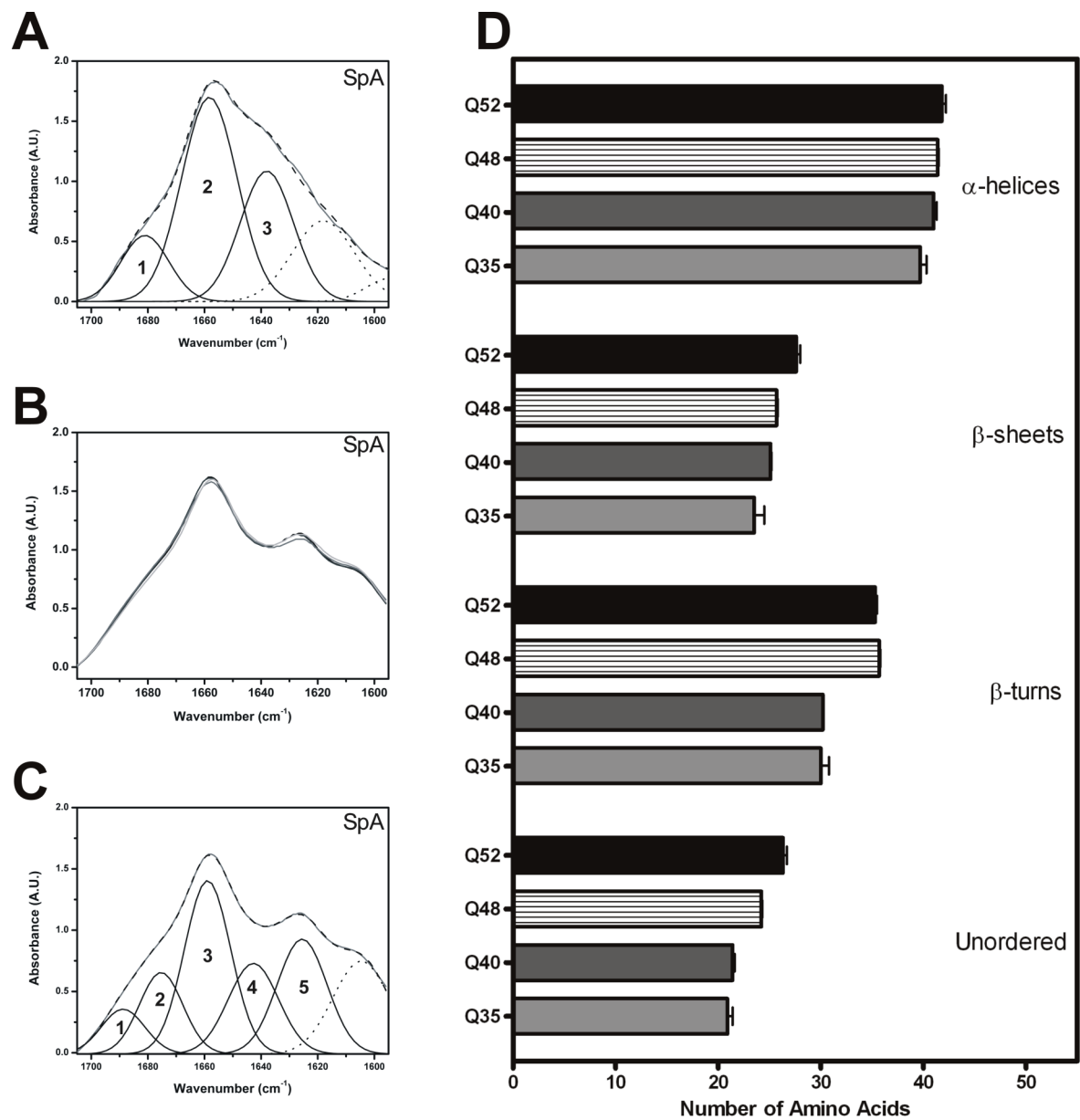


Figure 2

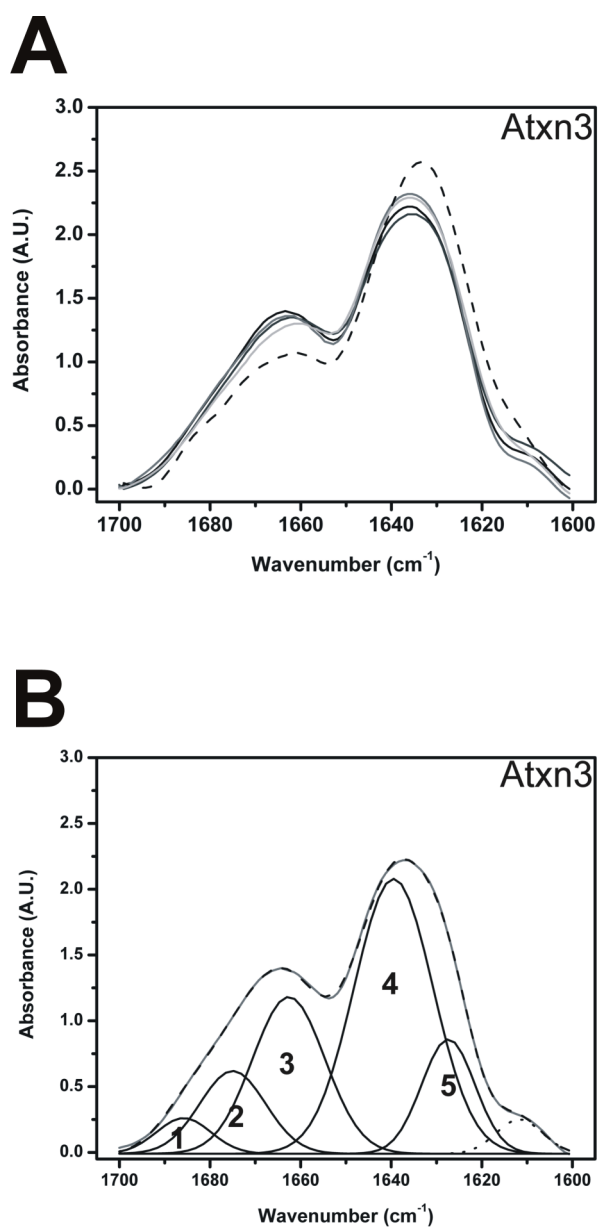
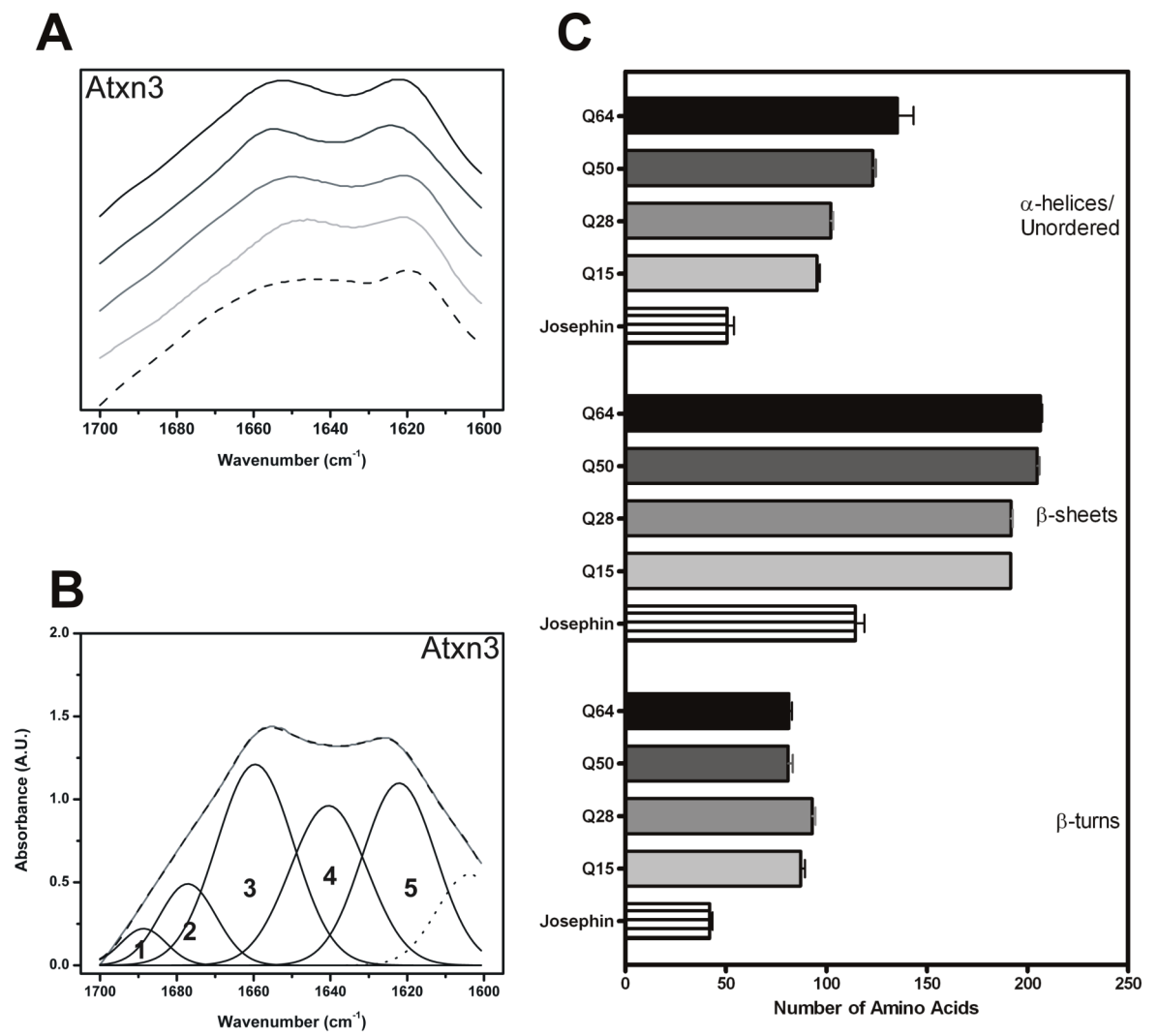


Figure 3





**Supplementary Data****Supplementary Table 1. Monomeric structure of protein A**

The secondary structure of SpA-cQ0 was determined via far-UV CD and FTIR, with similar results. The percentages of each structure present are expressed in terms of the mean (ave.) and standard deviation (S.D.). The means obtained by the two techniques are averaged and displayed in the far right column.

| Secondary Structure | CD       |      | FTIR     |      | Ave %       |
|---------------------|----------|------|----------|------|-------------|
|                     | Ave. (%) | S.D. | Ave. (%) | S.D. |             |
| $\alpha$ -helices   | 53.9     | 2.6  | 51.4     | 0.5  | <b>52.6</b> |
| $\beta$ -sheets     | 0        | 0.1  | 0.0      | 0.0  | <b>0</b>    |
| $\beta$ -turns      | 13.4     | 0.9  | 16.6     | 0.2  | <b>15.0</b> |
| Unordered           | 32.8     | 3    | 32.1     | 0.6  | <b>32.5</b> |

**Supplementary Table 2. Monomeric structure of ataxin-3 variants**

The secondary structure of the ataxin-3 variants were analysed via far-UV CD and FTIR, with similar results. The percentages of each structure present are expressed in terms of the mean (ave.) and standard deviation (S.D.). The means obtained by the two techniques are averaged and displayed in the far right column.

|                   | CD       |      | FTIR     |      |          |
|-------------------|----------|------|----------|------|----------|
| Josephin          | Ave. (%) | S.D. | Ave. (%) | S.D. | Ave. (%) |
| $\alpha$ -helices | 31.8     | 6.4  | 22.3     | 3.7  | 27.0     |
| $\beta$ -sheets   | 15.3     | 6.3  | 25.1     | 2.2  | 20.2     |
| $\beta$ -turns    | 16.8     | 1.7  | 12.8     | 1.1  | 14.8     |
| Unordered         | 36.2     | 2.9  | 39.8     | 2.8  | 38.0     |
| Atxn3(Q15)        | Ave.     | S.D. | Ave.     | S.D. | Ave.     |
| $\alpha$ -helices | 28.3     | 5.0  | 23.4     | 4.8  | 25.8     |
| $\beta$ -sheets   | 11.3     | 4.9  | 21.4     | 1.3  | 16.4     |
| $\beta$ -turns    | 15.4     | 2.1  | 11.6     | 2.5  | 13.5     |
| Unordered         | 45.0     | 4.1  | 43.6     | 3.2  | 44.3     |
| Atxn3(Q28)        | 43.9     | 0.4  | Ave.     | S.D. | Ave.     |
| $\alpha$ -helices | 25.3     | 4.7  | 22.1     | 1.0  | 23.7     |
| $\beta$ -sheets   | 15.5     | 3.8  | 20.8     | 1.5  | 18.2     |
| $\beta$ -turns    | 15.5     | 1.8  | 14.8     | 0.9  | 15.1     |
| Unordered         | 43.7     | 1.5  | 42.2     | 1.6  | 43.0     |
| Atxn3(Q50)        | Ave.     | S.D. | Ave.     | S.D. | Ave.     |
| $\alpha$ -helices | 27.7     | 5.5  | 23.4     | 0.4  | 25.5     |
| $\beta$ -sheets   | 11.0     | 5.0  | 18.7     | 0.6  | 14.9     |
| $\beta$ -turns    | 14.9     | 1.6  | 14.9     | 0.7  | 14.9     |
| Unordered         | 47.9     | 3.2  | 43.0     | 0.3  | 45.4     |
| Atxn3(Q64)        | Ave.     | S.D. | Ave.     | S.D. | Ave.     |
| $\alpha$ -helices | 27.6     | 6.3  | 23.7     | 0.6  | 25.6     |
| $\beta$ -sheets   | 11.0     | 7.0  | 18.2     | 0.4  | 14.6     |
| $\beta$ -turns    | 14.5     | 1.1  | 14.4     | 0.5  | 14.5     |
| Unordered         | 46.9     | 2.8  | 43.7     | 0.2  | 45.3     |

## **References**

## **References**

1. Chiti, F., and C. M. Dobson. 2006. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333-366.
2. Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. *Science* 181:223-230.
3. Jahn, T. R., and S. E. Radford. 2008. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys* 469:100-117.
4. Stefani, M., and C. M. Dobson. 2003. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* 81:678-699.
5. Balch, W. E., R. I. Morimoto, A. Dillin, and J. W. Kelly. 2008. Adapting proteostasis for disease intervention. *Science* 319:916-919.
6. Rambaran, R. N., and L. C. Serpell. 2008. Amyloid fibrils: abnormal protein assembly. *Prion* 2:112-117.
7. Wanker, E. E., E. Scherzinger, V. Heiser, A. Sittler, H. Eickhoff, and H. Lehrach. 1999. Membrane filter assay for detection of amyloid-like polyglutamine-containing protein aggregates. *Methods Enzymol* 309:375-386.
8. Groenning, M. 2009. Binding mode of Thioflavin T and other molecular probes in the context of amyloid fibrils-current status. *J Chem Biol*.
9. Kayed, R., E. Head, F. Sarsoza, T. Saing, C. W. Cotman, M. Necula, L. Margol, J. Wu, L. Breydo, J. L. Thompson, S. Rasool, T. Gurlo, P. Butler, and C. G. Glabe. 2007. Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. *Mol Neurodegener* 2:18.
10. O'Nuallain, B., and R. Wetzel. 2002. Conformational Abs recognizing a generic amyloid fibril epitope. *Proc Natl Acad Sci U S A* 99:1485-1490.
11. Nilsson, M. R. 2004. Techniques to study amyloid fibril formation in vitro. *Methods* 34:151-160.
12. Sunde, M., L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, and C. C. Blake. 1997. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* 273:729-739.
13. Stromer, T., and L. C. Serpell. 2005. Structure and morphology of the Alzheimer's amyloid fibril. *Microsc Res Tech* 67:210-217.
14. Makin, O. S., and L. C. Serpell. 2005. Structures for amyloid fibrils. *FEBS J* 272:5950-5961.
15. Eanes, E. D., and G. G. Glenner. 1968. X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem* 16:673-677.
16. Nelson, R., M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riek, R. Grothe, and D. Eisenberg. 2005. Structure of the cross-beta spine of amyloid-like fibrils. *Nature* 435:773-778.
17. Sawaya, M. R., S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riek, and D. Eisenberg. 2007. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* 447:453-457.
18. Sambashivan, S., Y. Liu, M. R. Sawaya, M. Gingery, and D. Eisenberg. 2005. Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature* 437:266-269.

19. Margittai, M., and R. Langen. 2008. Fibrils with parallel in-register structure constitute a major class of amyloid fibrils: molecular insights from electron paramagnetic resonance spectroscopy. *Q Rev Biophys* 41:265-297.
20. Wasmer, C., A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, and B. H. Meier. 2008. Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* 319:1523-1526.
21. Shewmaker, F., R. P. McGlinchey, K. R. Thurber, P. McPhie, F. Dyda, R. Tycko, and R. B. Wickner. 2009. The functional curli amyloid is not based on in-register parallel beta-sheet structure. *J Biol Chem* 284:25065-25076.
22. Petkova, A. T., Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko. 2002. A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* 99:16742-16747.
23. Jaroniec, C. P., C. E. MacPhee, N. S. Astrof, C. M. Dobson, and R. G. Griffin. 2002. Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proc Natl Acad Sci U S A* 99:16748-16753.
24. Jaroniec, C. P., C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson, and R. G. Griffin. 2004. High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc Natl Acad Sci U S A* 101:711-716.
25. Petkova, A. T., and R. Tycko. 2002. Sensitivity enhancement in structural measurements by solid state NMR through pulsed spin locking. *J Magn Reson* 155:293-299.
26. Tycko, R. 2006. Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q Rev Biophys* 39:1-55.
27. Fernandez-Escamilla, A. M., F. Rousseau, J. Schymkowitz, and L. Serrano. 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 22:1302-1306.
28. Tartaglia, G. G., A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, and M. Vendruscolo. 2008. Prediction of aggregation-prone regions in structured proteins. *J Mol Biol* 380:425-436.
29. Bryan, A. W., Jr., M. Menke, L. J. Cowen, S. L. Lindquist, and B. Berger. 2009. BETASCAN: probable beta-amyloids identified by pairwise probabilistic analysis. *PLoS Comput Biol* 5:e1000333.
30. Chiti, F., M. Stefani, N. Taddei, G. Ramponi, and C. M. Dobson. 2003. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424:805-808.
31. Wang, W., and M. H. Hecht. 2002. Rationally designed mutations convert de novo amyloid-like fibrils into monomeric beta-sheet proteins. *Proc Natl Acad Sci U S A* 99:2760-2765.
32. Makin, O. S., E. Atkins, P. Sikorski, J. Johansson, and L. C. Serpell. 2005. Molecular basis for amyloid fibril formation and stability. *Proc Natl Acad Sci U S A* 102:315-320.
33. Rousseau, F., J. Schymkowitz, and L. Serrano. 2006. Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* 16:118-126.
34. Oma, Y., Y. Kino, N. Sasagawa, and S. Ishiura. 2004. Intracellular localization of homopolymeric amino acid-containing proteins expressed in mammalian cells. *J Biol Chem* 279:21217-21222.
35. Oma, Y., Y. Kino, K. Toriumi, N. Sasagawa, and S. Ishiura. 2007. Interactions between homopolymeric amino acids (HPAAs). *Protein Sci* 16:2195-2204.

36. Bai, M., J. M. Zhou, and S. Perrett. 2004. The yeast prion protein Ure2 shows glutathione peroxidase activity in both native and fibrillar forms. *J Biol Chem* 279:50025-50030.
37. Krzewska, J., M. Tanaka, S. G. Burston, and R. Melki. 2007. Biochemical and functional analysis of the assembly of full-length Sup35p and its prion-forming domain. *J Biol Chem* 282:1679-1686.
38. Sackewitz, M., S. von Einem, G. Hause, M. Wunderlich, F. X. Schmid, and E. Schwarz. 2008. A folded and functional protein domain in an amyloid-like fibril. *Protein Sci* 17:1044-1054.
39. Staniforth, R. A., S. Giannini, L. D. Higgins, M. J. Conroy, A. M. Hounslow, R. Jerala, C. J. Craven, and J. P. Waltho. 2001. Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO J* 20:4774-4781.
40. Gras, S. L., A. K. Tickler, A. M. Squires, G. L. Devlin, M. A. Horton, C. M. Dobson, and C. E. MacPhee. 2008. Functionalised amyloid fibrils for roles in cell adhesion. *Biomaterials* 29:1553-1562.
41. Pawar, A. P., K. F. Dubay, J. Zurdo, F. Chiti, M. Vendruscolo, and C. M. Dobson. 2005. Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J Mol Biol* 350:379-392.
42. Chiti, F., and C. M. Dobson. 2009. Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5:15-22.
43. Monsellier, E., and F. Chiti. 2007. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep* 8:737-742.
44. Dumoulin, M., D. Canet, A. M. Last, E. Pardon, D. B. Archer, S. Muyldermans, L. Wyns, A. Matagne, C. V. Robinson, C. Redfield, and C. M. Dobson. 2005. Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations. *J Mol Biol* 346:773-788.
45. Booth, D. R., M. Sunde, V. Bellotti, C. V. Robinson, W. L. Hutchinson, P. E. Fraser, P. N. Hawkins, C. M. Dobson, S. E. Radford, C. C. Blake, and M. B. Pepys. 1997. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* 385:787-793.
46. Platt, G. W., V. J. McParland, A. P. Kalverda, S. W. Homans, and S. E. Radford. 2005. Dynamics in the unfolded state of beta2-microglobulin studied by NMR. *J Mol Biol* 346:279-294.
47. Bucciantini, M., E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, and M. Stefani. 2002. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* 416:507-511.
48. Guijarro, J. I., M. Sunde, J. A. Jones, I. D. Campbell, and C. M. Dobson. 1998. Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* 95:4224-4228.
49. Bucciantini, M., G. Calloni, F. Chiti, L. Formigli, D. Nosi, C. M. Dobson, and M. Stefani. 2004. Prefibrillar amyloid protein aggregates share common features of cytotoxicity. *J Biol Chem* 279:31374-31382.
50. Canet, D., A. M. Last, P. Tito, M. Sunde, A. Spencer, D. B. Archer, C. Redfield, C. V. Robinson, and C. M. Dobson. 2002. Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat Struct Biol* 9:308-315.

51. Chow, M. K., A. M. Ellisdon, L. D. Cabrita, and S. P. Bottomley. 2004. Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. *J Biol Chem* 279:47643-47651.
52. Plakoutsi, G., N. Taddei, M. Stefani, and F. Chiti. 2004. Aggregation of the Acylphosphatase from *Sulfolobus solfataricus*: the folded and partially unfolded states can both be precursors for amyloid formation. *J Biol Chem* 279:14111-14119.
53. Jahn, T. R., M. J. Parker, S. W. Homans, and S. E. Radford. 2006. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol* 13:195-201.
54. Elam, J. S., A. B. Taylor, R. Strange, S. Antonyuk, P. A. Doucette, J. A. Rodriguez, S. S. Hasnain, L. J. Hayward, J. S. Valentine, T. O. Yeates, and P. J. Hart. 2003. Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. *Nat Struct Biol* 10:461-467.
55. Bemporad, F., T. Vannocci, L. Varela, A. I. Azuaga, and F. Chiti. 2008. A model for the aggregation of the acylphosphatase from *Sulfolobus solfataricus* in its native-like state. *Biochim Biophys Acta* 1784:1986-1996.
56. Banci, L., I. Bertini, N. D'Amelio, E. Gaggelli, E. Libralesso, I. Matecko, P. Turano, and J. S. Valentine. 2005. Fully metallated S134N Cu,Zn-superoxide dismutase displays abnormal mobility and intermolecular contacts in solution. *J Biol Chem* 280:35815-35821.
57. Uversky, V. N. 2008. Amyloidogenesis of natively unfolded proteins. *Curr Alzheimer Res* 5:260-287.
58. Hook, V., I. Schechter, H. U. Demuth, and G. Hook. 2008. Alternative pathways for production of beta-amyloid peptides of Alzheimer's disease. *Biol Chem* 389:993-1006.
59. Calamai, M., N. Taddei, M. Stefani, G. Ramponi, and F. Chiti. 2003. Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins. *Biochemistry* 42:15078-15083.
60. Miyazaki, D., M. Yazaki, T. Gono, F. Kametani, A. Tsuchiya, M. Matsuda, Y. Takenaka, Y. Hosh, 2nd, and S. Ikeda. 2008. AH amyloidosis associated with an immunoglobulin heavy chain variable region (VH1) fragment: a case report. *Amyloid* 15:125-128.
61. Ferrone, F. 1999. Analysis of protein aggregation kinetics. *Methods Enzymol* 309:256-274.
62. Kheterpal, I., K. D. Cook, and R. Wetzel. 2006. Hydrogen/deuterium exchange mass spectrometry analysis of protein aggregates. *Methods Enzymol* 413:140-166.
63. Chen, S., F. A. Ferrone, and R. Wetzel. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci U S A* 99:11884-11889.
64. O'Nuallain, B., A. D. Williams, P. Westermarck, and R. Wetzel. 2004. Seeding specificity in amyloid growth induced by heterologous fibrils. *J Biol Chem* 279:17490-17499.
65. Cannon, M. J., A. D. Williams, R. Wetzel, and D. G. Myszka. 2004. Kinetic analysis of beta-amyloid fibril elongation. *Anal Biochem* 328:67-75.
66. Ellisdon, A. M., M. C. Pearce, and S. P. Bottomley. 2007. Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. *J Mol Biol* 368:595-605.

67. Harper, J. D., S. S. Wong, C. M. Lieber, and P. T. Lansbury. 1997. Observation of metastable Abeta amyloid protofibrils by atomic force microscopy. *Chem Biol* 4:119-125.
68. Kheterpal, I., M. Chen, K. D. Cook, and R. Wetzel. 2006. Structural differences in Abeta amyloid protofibrils and fibrils mapped by hydrogen exchange--mass spectrometry with on-line proteolytic fragmentation. *J Mol Biol* 361:785-795.
69. Myers, S. L., N. H. Thomson, S. E. Radford, and A. E. Ashcroft. 2006. Investigating the structural properties of amyloid-like fibrils formed in vitro from beta2-microglobulin using limited proteolysis and electrospray ionisation mass spectrometry. *Rapid Commun Mass Spectrom* 20:1628-1636.
70. Kaylor, J., N. Bodner, S. Edridge, G. Yamin, D. P. Hong, and A. L. Fink. 2005. Characterization of oligomeric intermediates in alpha-synuclein fibrillation: FRET studies of Y125W/Y133F/Y136F alpha-synuclein. *J Mol Biol* 353:357-372.
71. O'Sullivan, D. B., C. E. Jones, S. R. Abdelraheim, A. R. Thompsett, M. W. Brazier, H. Toms, D. R. Brown, and J. H. Viles. 2007. NMR characterization of the pH 4 beta-intermediate of the prion protein: the N-terminal half of the protein remains unstructured and retains a high degree of flexibility. *Biochem J* 401:533-540.
72. Yu, L., R. Edalji, J. E. Harlan, T. F. Holzman, A. P. Lopez, B. Labkovsky, H. Hillen, S. Barghorn, U. Ebert, P. L. Richardson, L. Miesbauer, L. Solomon, D. Bartley, K. Walter, R. W. Johnson, P. J. Hajduk, and E. T. Olejniczak. 2009. Structural characterization of a soluble amyloid beta-peptide oligomer. *Biochemistry* 48:1870-1877.
73. Huang, H., J. Milojevic, and G. Melacini. 2008. Analysis and optimization of saturation transfer difference NMR experiments designed to map early self-association events in amyloidogenic peptides. *J Phys Chem B* 112:5795-5802.
74. Orte, A., N. R. Birkett, R. W. Clarke, G. L. Devlin, C. M. Dobson, and D. Klenerman. 2008. Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc Natl Acad Sci U S A* 105:14424-14429.
75. Glabe, C. G. 2008. Structural classification of toxic amyloid oligomers. *J Biol Chem* 283:29639-29643.
76. Kayed, R., A. Pensalfini, L. Margol, Y. Sokolov, F. Sarsoza, E. Head, J. Hall, and C. Glabe. 2009. Annular protofibrils are a structurally and functionally distinct type of amyloid oligomer. *J Biol Chem* 284:4230-4237.
77. Kayed, R., E. Head, J. L. Thompson, T. M. McIntire, S. C. Milton, C. W. Cotman, and C. G. Glabe. 2003. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* 300:486-489.
78. Lashuel, H. A., and P. T. Lansbury, Jr. 2006. Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? *Q Rev Biophys* 39:167-201.
79. Ding, T. T., S. J. Lee, J. C. Rochet, and P. T. Lansbury, Jr. 2002. Annular alpha-synuclein protofibrils are produced when spherical protofibrils are incubated in solution or bound to brain-derived membranes. *Biochemistry* 41:10209-10217.
80. Terry, R. D. 1996. The pathogenesis of Alzheimer disease: an alternative to the amyloid hypothesis. *J Neuropathol Exp Neurol* 55:1023-1025.
81. Westerman, M. A., D. Cooper-Blacketer, A. Mariash, L. Kotilinek, T. Kawarabayashi, L. H. Younkin, G. A. Carlson, S. G. Younkin, and K. H. Ashe. 2002. The relationship between Abeta and memory in the Tg2576 mouse model of Alzheimer's disease. *J Neurosci* 22:1858-1867.



82. Caughey, B., and P. T. Lansbury. 2003. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* 26:267-298.
83. Hoshi, M., M. Sato, S. Matsumoto, A. Noguchi, K. Yasutake, N. Yoshida, and K. Sato. 2003. Spherical aggregates of beta-amyloid (amylospheroid) show high neurotoxicity and activate tau protein kinase I/glycogen synthase kinase-3beta. *Proc Natl Acad Sci U S A* 100:6370-6375.
84. Lambert, M. P., A. K. Barlow, B. A. Chromy, C. Edwards, R. Freed, M. Liosatos, T. E. Morgan, I. Rozovsky, B. Trommer, K. L. Viola, P. Wals, C. Zhang, C. E. Finch, G. A. Krafft, and W. L. Klein. 1998. Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. *Proc Natl Acad Sci U S A* 95:6448-6453.
85. Silveira, J. R., G. J. Raymond, A. G. Hughson, R. E. Race, V. L. Sim, S. F. Hayes, and B. Caughey. 2005. The most infectious prion protein particles. *Nature* 437:257-261.
86. Hyun, D. H., M. Lee, N. Hattori, S. Kubo, Y. Mizuno, B. Halliwell, and P. Jenner. 2002. Effect of wild-type or mutant Parkin on oxidative damage, nitric oxide, antioxidant defenses, and the proteasome. *J Biol Chem* 277:28572-28577.
87. Gusella, J. F., and M. E. MacDonald. 2000. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nat Rev Neurosci* 1:109-115.
88. Nance, M. A. 1997. Clinical aspects of CAG repeat diseases. *Brain Pathol* 7:881-900.
89. Ohshima, K., S. Kang, and R. D. Wells. 1996. CTG triplet repeats from human hereditary diseases are dominant genetic expansion products in *Escherichia coli*. *J Biol Chem* 271:1853-1856.
90. Sarkar, P. S., H. C. Chang, F. B. Boudi, and S. Reddy. 1998. CTG repeats show bimodal amplification in *E. coli*. *Cell* 95:531-540.
91. Wanker, E. E. 2000. Protein aggregation and pathogenesis of Huntington's disease: mechanisms and correlations. *Biol Chem* 381:937-942.
92. Ordway, J. M., S. Tallaksen-Greene, C. A. Gutekunst, E. M. Bernstein, J. A. Cearley, H. W. Wiener, L. S. t. Dure, R. Lindsey, S. M. Hersch, R. S. Jope, R. L. Albin, and P. J. Detloff. 1997. Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell* 91:753-763.
93. Ignatova, Z., and L. M. Gierasch. 2006. Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein. *J Biol Chem* 281:12959-12967.
94. Robertson, A. L., J. Horne, A. M. Ellisdon, B. Thomas, M. J. Scanlon, and S. P. Bottomley. 2008. The structural impact of a polyglutamine tract is location-dependent. *Biophys J* 95:5922-5930.
95. Borgia, A., P. M. Williams, and J. Clarke. 2008. Single-molecule studies of protein folding. *Annu Rev Biochem* 77:101-125.
96. Komure, O., A. Sano, N. Nishino, N. Yamauchi, S. Ueno, K. Kondoh, N. Sano, M. Takahashi, N. Murayama, I. Kondo, and et al. 1995. DNA analysis in hereditary dentatorubral-pallidoluyian atrophy: correlation between CAG repeat length and phenotypic variation and the molecular basis of anticipation. *Neurology* 45:143-149.
97. Nagafuchi, S., H. Yanagisawa, E. Ohsaki, T. Shirayama, K. Tadokoro, T. Inoue, and M. Yamada. 1994. Structure and expression of the gene responsible for the

- triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA). *Nat Genet* 8:177-182.
98. Kelley, N. W., X. Huang, S. Tam, C. Spiess, J. Frydman, and V. S. Pande. 2009. The predicted structure of the headpiece of the Huntingtin protein and its implications on Huntingtin aggregation. *J Mol Biol* 388:919-927.
  99. Li, W., L. C. Serpell, W. J. Carter, D. C. Rubinsztein, and J. A. Huntington. 2006. Expression and characterization of full-length human huntingtin, an elongated HEAT repeat protein. *J Biol Chem* 281:15916-15922.
  100. Zoppi, S., M. Marcelli, J. P. Deslypere, J. E. Griffin, J. D. Wilson, and M. J. McPhaul. 1992. Amino acid substitutions in the DNA-binding domain of the human androgen receptor are a frequent cause of receptor-binding positive androgen resistance. *Mol Endocrinol* 6:409-415.
  101. La Spada, A. R., E. M. Wilson, D. B. Lubahn, A. E. Harding, and K. H. Fischbeck. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352:77-79.
  102. Banfi, S., A. Servadio, M. Y. Chung, T. J. Kwiatkowski, Jr., A. E. McCall, L. A. Duvick, Y. Shen, E. J. Roth, H. T. Orr, and H. Y. Zoghbi. 1994. Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nat Genet* 7:513-520.
  103. Matilla, T., V. Volpini, D. Genis, J. Rosell, J. Corral, A. Davalos, A. Molins, and X. Estivill. 1993. Presymptomatic analysis of spinocerebellar ataxia type 1 (SCA1) via the expansion of the SCA1 CAG-repeat in a large pedigree displaying anticipation and parental male bias. *Hum Mol Genet* 2:2123-2128.
  104. Gispert, S., R. Twells, G. Orozco, A. Brice, J. Weber, L. Heredero, K. Scheufler, B. Riley, R. Allotey, C. Nothers, and et al. 1993. Chromosomal assignment of the second locus for autosomal dominant cerebellar ataxia (SCA2) to chromosome 12q23-24.1. *Nat Genet* 4:295-299.
  105. Pulst, S. M., N. Santos, D. Wang, H. Yang, D. Huynh, L. Velazquez, and K. P. Figueroa. 2005. Spinocerebellar ataxia type 2: polyQ repeat variation in the CACNA1A calcium channel modifies age of onset. *Brain* 128:2297-2303.
  106. Goto, J., M. Watanabe, Y. Ichikawa, S. B. Yee, N. Ihara, K. Endo, S. Igarashi, Y. Takiyama, C. Gaspar, P. Maciel, S. Tsuji, G. A. Rouleau, and I. Kanazawa. 1997. Machado-Joseph disease gene products carrying different carboxyl termini. *Neurosci Res* 28:373-377.
  107. Padiath, Q. S., A. K. Srivastava, S. Roy, S. Jain, and S. K. Brahmachari. 2005. Identification of a novel 45 repeat unstable allele associated with a disease phenotype at the MJD1/SCA3 locus. *Am J Med Genet B Neuropsychiatr Genet* 133B:124-126.
  108. Toru, S., T. Murakoshi, K. Ishikawa, H. Saegusa, H. Fujigasaki, T. Uchihara, S. Nagayama, M. Osanai, H. Mizusawa, and T. Tanabe. 2000. Spinocerebellar ataxia type 6 mutation alters P-type calcium channel function. *J Biol Chem* 275:10893-10898.
  109. Zhuchenko, O., J. Bailey, P. Bonnen, T. Ashizawa, D. W. Stockton, C. Amos, W. B. Dobyns, S. H. Subramony, H. Y. Zoghbi, and C. C. Lee. 1997. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat Genet* 15:62-69.
  110. Benomar, A., L. Krols, G. Stevanin, G. Cancel, E. LeGuern, G. David, H. Ouhabi, J. J. Martin, A. Durr, A. Zaim, and et al. 1995. The gene for autosomal

- 
- dominant cerebellar ataxia with pigmentary macular dystrophy maps to chromosome 3p12-p21.1. *Nat Genet* 10:84-88.
111. David, G., N. Abbas, G. Stevanin, A. Durr, G. Yvert, G. Cancel, C. Weber, G. Imbert, F. Saudou, E. Antoniou, H. Drabkin, R. Gemmill, P. Giunti, A. Benomar, N. Wood, M. Ruberg, Y. Agid, J. L. Mandel, and A. Brice. 1997. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* 17:65-70.
  112. Peterson, M. G., N. Tanese, B. F. Pugh, and R. Tjian. 1990. Functional domains and upstream activation properties of cloned human TATA binding protein. *Science* 248:1625-1630.
  113. Nakamura, K., S. Y. Jeong, T. Uchihara, M. Anno, K. Nagashima, T. Nagashima, S. Ikeda, S. Tsuji, and I. Kanazawa. 2001. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet* 10:1441-1448.
  114. Chow, M. K., A. M. Ellisdon, L. D. Cabrita, and S. P. Bottomley. 2006. Purification of polyglutamine proteins. *Methods Enzymol* 413:1-19.
  115. Perutz, M. F., T. Johnson, M. Suzuki, and J. T. Finch. 1994. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A* 91:5355-5358.
  116. Chen, S., V. Berthelie, W. Yang, and R. Wetzel. 2001. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J Mol Biol* 311:173-182.
  117. Nagai, Y., T. Inui, H. A. Popiel, N. Fujikake, K. Hasegawa, Y. Urade, Y. Goto, H. Naiki, and T. Toda. 2007. A toxic monomeric conformer of the polyglutamine protein. *Nat Struct Mol Biol* 14:332-340.
  118. Nagai, Y., T. Tucker, H. Ren, D. J. Kenan, B. S. Henderson, J. D. Keene, W. J. Strittmatter, and J. R. Burke. 2000. Inhibition of polyglutamine protein aggregation and cell death by novel peptides identified by phage display screening. *J Biol Chem* 275:10437-10442.
  119. Esposito, L., A. Paladino, C. Pedone, and L. Vitagliano. 2008. Insights into structure, stability, and toxicity of monomeric and aggregated polyglutamine models from molecular dynamics simulations. *Biophys J* 94:4031-4040.
  120. Ogawa, H., M. Nakano, H. Watanabe, E. B. Starikov, S. M. Rothstein, and S. Tanaka. 2008. Molecular dynamics simulation study on the structural stabilities of polyglutamine peptides. *Comput Biol Chem* 32:102-110.
  121. Marchut, A. J., and C. K. Hall. 2007. Effects of chain length on the aggregation of model polyglutamine peptides: molecular dynamics simulations. *Proteins* 66:96-109.
  122. Klein, F. A., A. Pastore, L. Masino, G. Zeder-Lutz, H. Nierengarten, M. Oulad-Abdelghani, D. Altschuh, J. L. Mandel, and Y. Trottier. 2007. Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. *J Mol Biol* 371:235-244.
  123. Bennett, M. J., K. E. Huey-Tubman, A. B. Herr, A. P. West, Jr., S. A. Ross, and P. J. Bjorkman. 2002. Inaugural Article: A linear lattice model for polyglutamine in CAG-expansion diseases. *Proc Natl Acad Sci U S A* 99:11634-11639.
  124. Li, P., K. E. Huey-Tubman, T. Gao, X. Li, A. P. West, Jr., M. J. Bennett, and P. J. Bjorkman. 2007. The structure of a polyQ-anti-polyQ complex reveals binding according to a linear lattice model. *Nat Struct Mol Biol* 14:381-387.

125. Wang, X., A. Vitalis, M. A. Wyczalkowski, and R. V. Pappu. 2006. Characterizing the conformational ensemble of monomeric polyglutamine. *Proteins* 63:297-311.
126. Mukhopadhyay, S., R. Krishnan, E. A. Lemke, S. Lindquist, and A. A. Deniz. 2007. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc Natl Acad Sci U S A* 104:2649-2654.
127. Uversky, V. N. 2003. A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. *J Biomol Struct Dyn* 21:211-234.
128. Nozaki, K., O. Onodera, H. Takano, and S. Tsuji. 2001. Amino acid sequences flanking polyglutamine stretches influence their potential for aggregate formation. *Neuroreport* 12:3357-3364.
129. Bhattacharyya, A., A. K. Thakur, V. M. Chellgren, G. Thiagarajan, A. D. Williams, B. W. Chellgren, T. P. Creamer, and R. Wetzel. 2006. Oligoproline effects on polyglutamine conformation and aggregation. *J Mol Biol* 355:524-535.
130. Duennwald, M. L., S. Jagadish, P. J. Muchowski, and S. Lindquist. 2006. Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc Natl Acad Sci U S A* 103:11045-11050.
131. Steward, A., S. Adhya, and J. Clarke. 2002. Sequence conservation in Ig-like domains: the role of highly conserved proline residues in the fibronectin type III superfamily. *J Mol Biol* 318:935-940.
132. Thakur, A. K., M. Jayaraman, R. Mishra, M. Thakur, V. M. Chellgren, L. B. II, D. H. Anjum, R. Kodali, T. P. Creamer, J. F. Conway, M. G. A., and R. Wetzel. 2009. Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat Struct Mol Biol*.
133. Sharma, D., S. Sharma, S. Pasha, and S. K. Brahmachari. 1999. Peptide models for inherited neurodegenerative disorders: conformation and aggregation properties of long polyglutamine peptides with and without interruptions. *FEBS Lett* 456:181-185.
134. Popiel, H. A., Y. Nagai, O. Onodera, T. Inui, N. Fujikake, Y. Urade, W. J. Strittmatter, J. R. Burke, A. Ichikawa, and T. Toda. 2004. Disruption of the toxic conformation of the expanded polyglutamine stretch leads to suppression of aggregate formation and cytotoxicity. *Biochem Biophys Res Commun* 317:1200-1206.
135. Thakur, A. K., and R. Wetzel. 2002. Mutational analysis of the structural organization of polyglutamine aggregates. *Proc Natl Acad Sci U S A* 99:17014-17019.
136. Busch, A., S. Engemann, R. Lurz, H. Okazawa, H. Lehrach, and E. E. Wanker. 2003. Mutant huntingtin promotes the fibrillogenesis of wild-type huntingtin: a potential mechanism for loss of huntingtin function in Huntington's disease. *J Biol Chem* 278:41452-41461.
137. Bevivino, A. E., and P. J. Loll. 2001. An expanded glutamine repeat destabilizes native ataxin-3 structure and mediates formation of parallel beta -fibrils. *Proc Natl Acad Sci U S A* 98:11955-11960.
138. Tartaglia, G. G., S. Pechmann, C. M. Dobson, and M. Vendruscolo. 2007. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci* 32:204-206.

139. Rousseau, F., L. Serrano, and J. W. Schymkowitz. 2006. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* 355:1037-1047.
140. Faux, N. G., S. P. Bottomley, A. M. Lesk, J. A. Irving, J. R. Morrison, M. G. de la Banda, and J. C. Whisstock. 2005. Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res* 15:537-551.
141. Tanaka, M., I. Morishima, T. Akagi, T. Hashikawa, and N. Nukina. 2001. Intra- and intermolecular beta-pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases. *J Biol Chem* 276:45470-45475.
142. Davies, P., K. Watt, S. M. Kelly, C. Clark, N. C. Price, and I. J. McEwan. 2008. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *J Mol Endocrinol* 41:301-314.
143. Chow, M. K., J. P. Mackay, J. C. Whisstock, M. J. Scanlon, and S. P. Bottomley. 2004. Structural and functional analysis of the Josephin domain of the polyglutamine protein ataxin-3. *Biochem Biophys Res Commun* 322:387-394.
144. Ladurner, A. G., and A. R. Fersht. 1997. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J Mol Biol* 273:330-337.
145. Wellington, C. L., L. M. Ellerby, A. S. Hackam, R. L. Margolis, M. A. Trifiro, R. Singaraja, K. McCutcheon, G. S. Salvesen, S. S. Propp, M. Bromm, K. J. Rowland, T. Zhang, D. Rasper, S. Roy, N. Thornberry, L. Pinsky, A. Kakizuka, C. A. Ross, D. W. Nicholson, D. E. Bredesen, and M. R. Hayden. 1998. Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. *J Biol Chem* 273:9158-9167.
146. Goti, D., S. M. Katzen, J. Mez, N. Kurtis, J. Kiluk, L. Ben-Haiem, N. A. Jenkins, N. G. Copeland, A. Kakizuka, A. H. Sharp, C. A. Ross, P. R. Mouton, and V. Colomer. 2004. A mutant ataxin-3 putative-cleavage fragment in brains of Machado-Joseph disease patients and transgenic mice is cytotoxic above a critical concentration. *J Neurosci* 24:10266-10279.
147. Nucifora, F. C., Jr., L. M. Ellerby, C. L. Wellington, J. D. Wood, W. J. Herring, A. Sawa, M. R. Hayden, V. L. Dawson, T. M. Dawson, and C. A. Ross. 2003. Nuclear localization of a non-caspase truncation product of atrophin-1, with an expanded polyglutamine repeat, increases cellular toxicity. *J Biol Chem* 278:13047-13055.
148. DiFiglia, M., E. Sapp, K. O. Chase, S. W. Davies, G. P. Bates, J. P. Vonsattel, and N. Aronin. 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* 277:1990-1993.
149. Ellerby, L. M., A. S. Hackam, S. S. Propp, H. M. Ellerby, S. Rabizadeh, N. R. Cashman, M. A. Trifiro, L. Pinsky, C. L. Wellington, G. S. Salvesen, M. R. Hayden, and D. E. Bredesen. 1999. Kennedy's disease: caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. *J Neurochem* 72:185-195.
150. Goldberg, Y. P., D. W. Nicholson, D. M. Rasper, M. A. Kalchman, H. B. Koide, R. K. Graham, M. Bromm, P. Kazemi-Esfarjani, N. A. Thornberry, J. P. Vaillancourt, and M. R. Hayden. 1996. Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nat Genet* 13:442-449.

151. Ellisdon, A. M., B. Thomas, and S. P. Bottomley. 2006. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J Biol Chem* 281:16888-16896.
152. Gales, L., L. Cortes, C. Almeida, C. V. Melo, M. do Carmo Costa, P. Maciel, D. T. Clarke, A. M. Damas, and S. Macedo-Ribeiro. 2005. Towards a structural understanding of the fibrillization pathway in Machado-Joseph's disease: trapping early oligomers of non-expanded ataxin-3. *J Mol Biol* 353:642-654.
153. Menon, R. P., and A. Pastore. 2006. Expansion of amino acid homo-sequences in proteins: insights into the role of amino acid homo-polymers and of the protein context in aggregation. *Cell Mol Life Sci* 63:1677-1685.
154. Atwal, R. S., J. Xia, D. Pinchev, J. Taylor, R. M. Epan, and R. Truant. 2007. Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. *Hum Mol Genet* 16:2600-2615.
155. Colby, D. W., Y. Chu, J. P. Cassady, M. Duennwald, H. Zazulak, J. M. Webster, A. Messer, S. Lindquist, V. M. Ingram, and K. D. Wittrup. 2004. Potent inhibition of huntingtin aggregation and cytotoxicity by a disulfide bond-free single-domain intracellular antibody. *Proc Natl Acad Sci U S A* 101:17616-17621.
156. Ignatova, Z., A. K. Thakur, R. Wetzel, and L. M. Gierasch. 2007. In-cell aggregation of a polyglutamine-containing chimera is a multistep process initiated by the flanking sequence. *J Biol Chem* 282:36736-36743.
157. de Chiara, C., R. P. Menon, S. Adinolfi, J. de Boer, E. Ktistaki, G. Kelly, L. Calder, D. Kioussis, and A. Pastore. 2005. The AXH domain adopts alternative folds the solution structure of HBP1 AXH. *Structure* 13:743-753.
158. Burright, E. N., J. D. Davidson, L. A. Duvick, B. Koshy, H. Y. Zoghbi, and H. T. Orr. 1997. Identification of a self-association region within the SCA1 gene product, ataxin-1. *Hum Mol Genet* 6:513-518.
159. de Chiara, C., C. Giannini, S. Adinolfi, J. de Boer, S. Guida, A. Ramos, C. Jodice, D. Kioussis, and A. Pastore. 2003. The AXH module: an independently folded domain common to ataxin-1 and HBP1. *FEBS Lett* 551:107-112.
160. Hong, S., S. J. Kim, S. Ka, I. Choi, and S. Kang. 2002. USP7, a ubiquitin-specific protease, interacts with ataxin-1, the SCA1 gene product. *Mol Cell Neurosci* 20:298-306.
161. Matilla, A., B. T. Koshy, C. J. Cummings, T. Isobe, H. T. Orr, and H. Y. Zoghbi. 1997. The cerebellar leucine-rich acidic nuclear protein interacts with ataxin-1. *Nature* 389:974-978.
162. Yue, S., H. G. Serra, H. Y. Zoghbi, and H. T. Orr. 2001. The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. *Hum Mol Genet* 10:25-30.
163. de Chiara, C., R. P. Menon, F. Dal Piaz, L. Calder, and A. Pastore. 2005. Polyglutamine is not all: the functional role of the AXH domain in the ataxin-1 protein. *J Mol Biol* 354:883-893.
164. Johansson, J., T. E. Weaver, and L. O. Tjernberg. 2004. Proteolytic generation and aggregation of peptides from transmembrane regions: lung surfactant protein C and amyloid beta-peptide. *Cell Mol Life Sci* 61:326-335.
165. Nicastro, G., R. P. Menon, L. Masino, P. P. Knowles, N. Q. McDonald, and A. Pastore. 2005. The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. *Proc Natl Acad Sci U S A* 102:10493-10498.

166. Chen, Y. W., M. D. Allen, D. B. Veprintsev, J. Lowe, and M. Bycroft. 2004. The structure of the AXH domain of spinocerebellar ataxin-1. *J Biol Chem* 279:3758-3765.
167. Bulone, D., L. Masino, D. J. Thomas, P. L. San Biagio, and A. Pastore. 2006. The interplay between PolyQ and protein context delays aggregation by forming a reservoir of protofibrils. *PLoS ONE* 1:e111.
168. Klein, A. F., M. Ebihara, C. Alexander, M. J. Dicaire, A. M. Sasseville, Y. Langelier, G. A. Rouleau, and B. Brais. 2008. PABPN1 polyalanine tract deletion and long expansions modify its aggregation pattern and expression. *Exp Cell Res* 314:1652-1666.
169. Fernandez-Bellot, E., E. Guillemet, A. Baudin-Baillieu, S. Gaumer, A. A. Komar, and C. Cullin. 1999. Characterization of the interaction domains of Ure2p, a prion-like protein of yeast. *Biochem J* 338 ( Pt 2):403-407.
170. Bilen, J., and N. M. Bonini. 2007. Genome-wide screen for modifiers of ataxin-3 neurodegeneration in *Drosophila*. *PLoS Genet* 3:1950-1964.
171. Carra, S., M. Sivilotti, A. T. Chavez Zobel, H. Lambert, and J. Landry. 2005. HspB8, a small heat shock protein mutated in human neuromuscular disorders, has in vivo chaperone activity in cultured cells. *Hum Mol Genet* 14:1659-1669.
172. Rudiger, S., A. Buchberger, and B. Bukau. 1997. Interaction of Hsp70 chaperones with substrates. *Nat Struct Biol* 4:342-349.
173. Muchowski, P. J., G. Schaffar, A. Sittler, E. E. Wanker, M. K. Hayer-Hartl, and F. U. Hartl. 2000. Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc Natl Acad Sci U S A* 97:7841-7846.
174. Wacker, J. L., M. H. Zareie, H. Fong, M. Sarikaya, and P. J. Muchowski. 2004. Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer. *Nat Struct Mol Biol* 11:1215-1222.
175. Kitamura, A., H. Kubota, C. G. Pack, G. Matsumoto, S. Hirayama, Y. Takahashi, H. Kimura, M. Kinjo, R. I. Morimoto, and K. Nagata. 2006. Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state. *Nat Cell Biol* 8:1163-1170.
176. Tam, S., R. Geller, C. Spiess, and J. Frydman. 2006. The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. *Nat Cell Biol* 8:1155-1162.
177. Behrends, C., C. A. Langer, R. Boteva, U. M. Bottcher, M. J. Stemp, G. Schaffar, B. V. Rao, A. Giese, H. Kretzschmar, K. Siegers, and F. U. Hartl. 2006. Chaperonin TRiC promotes the assembly of polyQ expansion proteins into nontoxic oligomers. *Mol Cell* 23:887-897.
178. Nollen, E. A., S. M. Garcia, G. van Haaften, S. Kim, A. Chavez, R. I. Morimoto, and R. H. Plasterk. 2004. Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation. *Proc Natl Acad Sci U S A* 101:6403-6408.
179. Branco, J., I. Al-Ramahi, L. Ukani, A. M. Perez, P. Fernandez-Funez, D. Rincon-Limas, and J. Botas. 2008. Comparative analysis of genetic modifiers in *Drosophila* points to common and distinct mechanisms of pathogenesis among polyglutamine diseases. *Hum Mol Genet* 17:376-390.
180. Iwata, A., Y. Nagashima, L. Matsumoto, T. Suzuki, T. Yamanaka, H. Date, K. Deoka, N. Nukina, and S. Tsuji. 2009. Intranuclear degradation of polyglutamine aggregates by the ubiquitin-proteasome system. *J Biol Chem* 284:9796-9803.

181. Ciechanover, A., and P. Brundin. 2003. The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron* 40:427-446.
182. Venkatraman, P., R. Wetzel, M. Tanaka, N. Nukina, and A. L. Goldberg. 2004. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol Cell* 14:95-104.
183. Bence, N. F., R. M. Sampat, and R. R. Kopito. 2001. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* 292:1552-1555.
184. Jana, N. R., E. A. Zemskov, G. Wang, and N. Nukina. 2001. Altered proteasomal function due to the expression of polyglutamine-expanded truncated N-terminal huntingtin induces apoptosis by caspase activation through mitochondrial cytochrome c release. *Hum Mol Genet* 10:1049-1059.
185. Chai, Y., L. Wu, J. D. Griffin, and H. L. Paulson. 2001. The role of protein composition in specifying nuclear inclusion formation in polyglutamine disease. *J Biol Chem* 276:44889-44897.
186. Brignull, H. R., J. F. Morley, and R. I. Morimoto. 2007. The stress of misfolded proteins: *C. elegans* models for neurodegenerative disease and aging. *Adv Exp Med Biol* 594:167-189.
187. Stenoien, D. L., M. Mielke, and M. A. Mancini. 2002. Intranuclear ataxin1 inclusions contain both fast- and slow-exchanging components. *Nat Cell Biol* 4:806-810.
188. Chai, Y., J. Shao, V. M. Miller, A. Williams, and H. L. Paulson. 2002. Live-cell imaging reveals divergent intracellular dynamics of polyglutamine disease proteins and supports a sequestration model of pathogenesis. *Proc Natl Acad Sci U S A* 99:9310-9315.
189. van Roon-Mom, W. M., S. J. Reid, R. L. Faull, and R. G. Snell. 2005. TATA-binding protein in neurodegenerative disease. *Neuroscience* 133:863-872.
190. Poletti, A. 2004. The polyglutamine tract of androgen receptor: from functions to dysfunctions in motor neurons. *Front Neuroendocrinol* 25:1-26.
191. Matilla, A., E. D. Roberson, S. Banfi, J. Morales, D. L. Armstrong, E. N. Burright, H. T. Orr, J. D. Sweatt, H. Y. Zoghbi, and M. M. Matzuk. 1998. Mice lacking ataxin-1 display learning deficits and decreased hippocampal paired-pulse facilitation. *J Neurosci* 18:5508-5516.
192. Friedman, M. J., A. G. Shah, Z. H. Fang, E. G. Ward, S. T. Warren, S. Li, and X. J. Li. 2007. Polyglutamine domain modulates the TBP-TFIIB interaction: implications for its normal function and neurodegeneration. *Nat Neurosci* 10:1519-1528.
193. Young, J. E., L. Gouw, S. Propp, B. L. Sopher, J. Taylor, A. Lin, E. Hermel, A. Logvinova, S. F. Chen, S. Chen, D. E. Bredesen, R. Truant, L. J. Ptacek, A. R. La Spada, and L. M. Ellerby. 2007. Proteolytic cleavage of ataxin-7 by caspase-7 modulates cellular toxicity and transcriptional dysregulation. *J Biol Chem* 282:30150-30160.
194. Helmlinger, D., S. Hardy, S. Sasorith, F. Klein, F. Robert, C. Weber, L. Miguet, N. Potier, A. Van-Dorsselaer, J. M. Wurtz, J. L. Mandel, L. Tora, and D. Devys. 2004. Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Hum Mol Genet* 13:1257-1265.
195. Pace, C. N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266-280.



196. Cabrita, L. D., W. Dai, and S. P. Bottomley. 2006. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol* 6:12.
197. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166:557-580.
198. Studier, F. W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41:207-234.
199. Marley, J., M. Lu, and C. Bracken. 2001. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* 20:71-75.
200. Gill, S. C., and P. H. von Hippel. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182:319-326.
201. Pace, C. N., F. Vajdos, L. Fee, G. Grimsley, and T. Gray. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4:2411-2423.
202. Andrade, M. A., P. Chacon, J. J. Merelo, and F. Moran. 1993. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein engineering* 6:383-390.
203. Lobley, A., L. Whitmore, and B. A. Wallace. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 18:211-212.
204. Santoro, M. M., and D. W. Bolen. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27:8063-8068.
205. Sattler, M., J. Schleucher, and C. Griesinger. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* 34:93-158.
206. Kay, L. E., P. Keifer, and P. Saarinen. 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* 114:10663-10665.
207. Hoshino, M., H. Katou, Y. Hagihara, K. Hasegawa, H. Naiki, and Y. Goto. 2002. Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange. *Nat Struct Biol* 9:332-336.
208. Ippel, J. H., A. Olofsson, J. Schleucher, E. Lundgren, and S. S. Wijmenga. 2002. Probing solvent accessibility of amyloid fibrils by solution NMR spectroscopy. *Proc Natl Acad Sci U S A* 99:8648-8653.
209. Monti, M., A. Amoresano, S. Giorgetti, V. Bellotti, and P. Pucci. 2005. Limited proteolysis in the investigation of beta2-microglobulin amyloidogenic and fibrillar states. *Biochim Biophys Acta* 1753:44-50.
210. Frare, E., M. F. Mossuto, P. Polverino de Laureto, M. Dumoulin, C. M. Dobson, and A. Fontana. 2006. Identification of the core structure of lysozyme amyloid fibrils by proteolysis. *J Mol Biol* 361:551-561.
211. Hall, D., N. Hirota, and C. M. Dobson. 2005. A toy model for predicting the rate of amyloid formation from unfolded protein. *J Mol Biol* 351:195-205.
212. Kim, M. W., Y. Chelliah, S. W. Kim, Z. Otwinowski, and I. Bezprozvanny. 2009. Secondary structure of Huntingtin amino-terminal region. *Structure* 17:1205-1212.
213. McMahon, S. J., M. G. Pray-Grant, D. Schieltz, J. R. Yates, 3rd, and P. A. Grant. 2005. Polyglutamine-expanded spinocerebellar ataxia-7 protein disrupts

- normal SAGA and SLIK histone acetyltransferase activity. *Proc Natl Acad Sci U S A* 102:8478-8482.
214. Hoffman, A., E. Sinn, T. Yamamoto, J. Wang, A. Roy, M. Horikoshi, and R. G. Roeder. 1990. Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature* 346:387-390.
  215. Stenoien, D. L., C. J. Cummings, H. P. Adams, M. G. Mancini, K. Patel, G. N. DeMartino, M. Marcelli, N. L. Weigel, and M. A. Mancini. 1999. Polyglutamine-expanded androgen receptors form aggregates that sequester heat shock proteins, proteasome components and SRC-1, and are suppressed by the HDJ-2 chaperone. *Hum Mol Genet* 8:731-741.
  216. Margolis, R. L., M. R. Abraham, S. B. Gatchell, S. H. Li, A. S. Kidwai, T. S. Breschel, O. C. Stine, C. Callahan, M. G. McInnis, and C. A. Ross. 1997. cDNAs with long CAG trinucleotide repeats from human brain. *Hum Genet* 100:114-122.
  217. McParland, V. J., N. M. Kad, A. P. Kalverda, A. Brown, P. Kirwin-Jones, M. G. Hunter, M. Sunde, and S. E. Radford. 2000. Partially unfolded states of beta(2)-microglobulin and amyloid formation in vitro. *Biochemistry* 39:8735-8746.
  218. Zhou, Z., J. B. Fan, H. L. Zhu, F. Shewmaker, X. Yan, X. Chen, J. Chen, G. F. Xiao, L. Guo, and Y. Liang. 2009. Crowded, Cell-like Environment Accelerates the Nucleation Step of Amyloidogenic Protein Misfolding. *J Biol Chem*.
  219. Smith, D. P., D. J. Tew, A. F. Hill, S. P. Bottomley, C. L. Masters, K. J. Barnham, and R. Cappai. 2008. Formation of a high affinity lipid-binding intermediate during the early aggregation phase of alpha-synuclein. *Biochemistry* 47:1425-1434.
  220. Ecroyd, H., and J. A. Carver. 2009. Crystallin proteins and amyloid fibrils. *Cell Mol Life Sci* 66:62-81.
  221. Haslbeck, M., T. Franzmann, D. Weinfurtner, and J. Buchner. 2005. Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol* 12:842-846.
  222. Vitalis, A., N. Lyle, and R. V. Pappu. 2009. Thermodynamics of beta-sheet formation in polyglutamine. *Biophys J* 97:303-311.
  223. Scherzinger, E., R. Lurz, M. Turmaine, L. Mangiarini, B. Hollenbach, R. Hasenbank, G. P. Bates, S. W. Davies, H. Lehrach, and E. E. Wanker. 1997. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* 90:549-558.
  224. Scherzinger, E., A. Sittler, K. Schweiger, V. Heiser, R. Lurz, R. Hasenbank, G. P. Bates, H. Lehrach, and E. E. Wanker. 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.
  225. Tanaka, M., Y. Machida, and N. Nukina. 2005. A novel therapeutic strategy for polyglutamine diseases by stabilizing aggregation-prone proteins with small molecules. *J Mol Med* 83:343-352.
  226. Ehrnhoefer, D. E., M. Duennwald, P. Markovic, J. L. Wacker, S. Engemann, M. Roark, J. Legleiter, J. L. Marsh, L. M. Thompson, S. Lindquist, P. J. Muchowski, and E. E. Wanker. 2006. Green tea (-)-epigallocatechin-gallate modulates early events in huntingtin misfolding and reduces toxicity in Huntington's disease models. *Hum Mol Genet* 15:2743-2751.