View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

The Structural Impact of a Polyglutamine Tract Is Location-Dependent

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

*Department of Biochemistry and Molecular Biology, Monash University, Clayton, Australia; and †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 β-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 ~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 monomeric 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 β-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 myglobin 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 β-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}N only or $^{15}N/^{13}C$ 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 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 μM 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 phenylmethylsulphonyl 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 μM. 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 μM 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 μM ThT, and fluorescence emission intensity at 490 nm was recorded ($\lambda ex = 450$) 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% Me_2SO/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 μM. Control assays with identical final Me_2SO concentrations were performed and showed no significant effect of Me_2SO addition.

Membrane filter trap assay

This assay was carried out with minor modifications to the previously described protocol (31). A 10 μL aliquot of protein 20 μM 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 μL with 2% (w/v) SDS. The samples were filtered through a cellulose acetate membrane (Schleicher and Schuell, 0.2 μm) using a Bio-Rad Bio-Dot SF microfiltration unit. After application, the membrane was washed twice by filtering through 150 μL 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 μM 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 μM. 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 μM 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 ϕ/ψ angle restraints were calculated based on chemical shifts for Cα, Cβ, Hα, 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 $> 5°$ 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 ≥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 α-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.

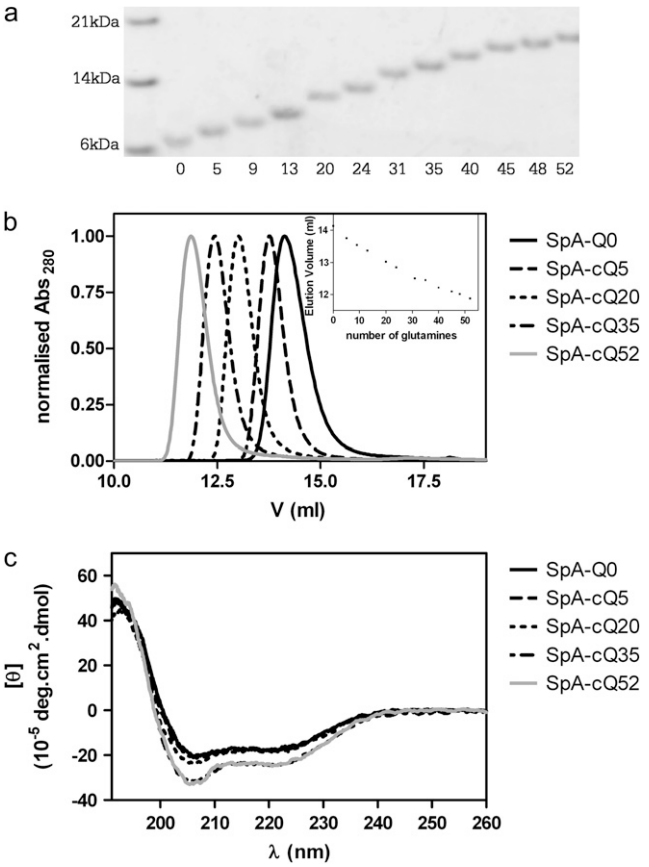

FIGURE 1 SpA-polyQ proteins are monomeric and the polyQ region is unstructured. (*a*) Purified SpA-polyQ fusion proteins were analyzed by 12% (w/v) SDS-PAGE. (*b*) SpA-polyQ fusion proteins were analyzed by SEC using a Superdex 75 column. The elution profiles of SpA-Q0, SpA-cQ5, SpA-cQ20, SpA-cQ35, and 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 ~270 h for SpA-cQ35 and decreasing exponentially to plateau at ~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-

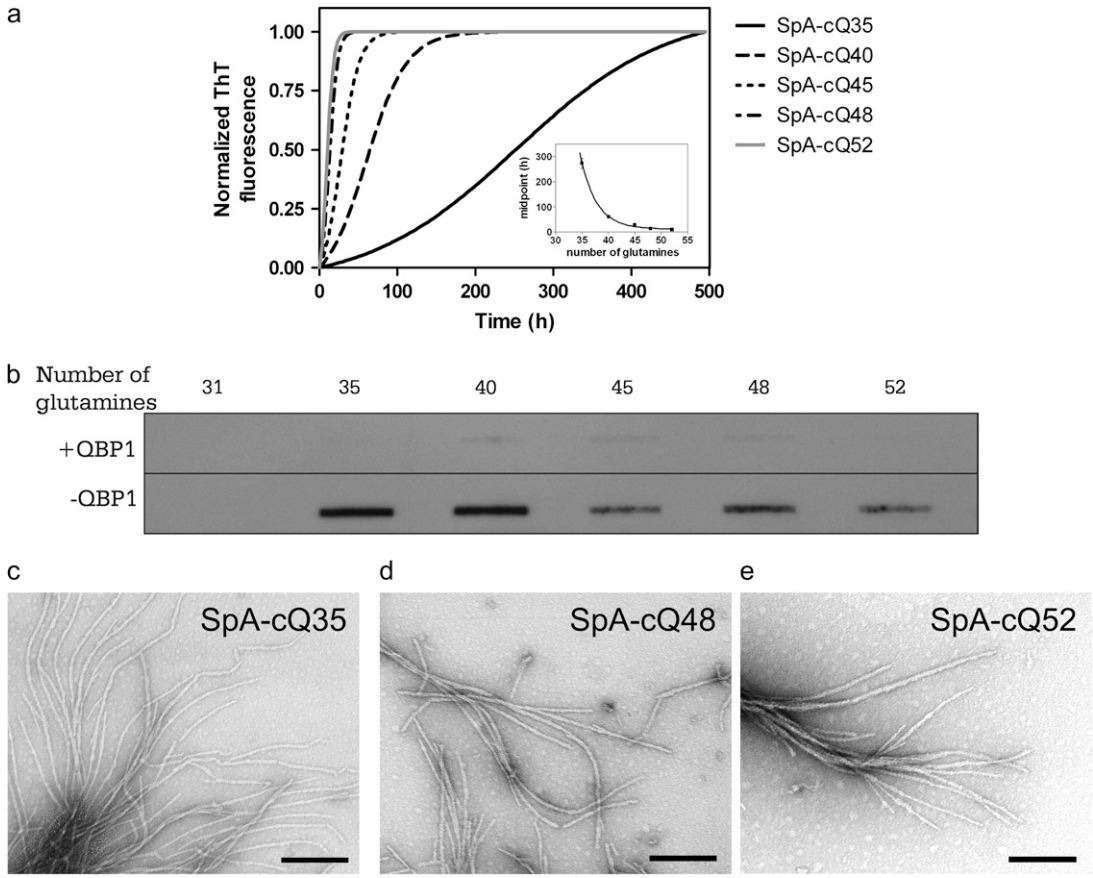

FIGURE 2 Expanded PolyQ proteins form typical fibrils. (*a*) ThT analysis of aggregation formation for proteins with repeat lengths ≥35. Proteins with repeat lengths below 35 did not aggregate over the 500 h timespan analyzed. Curves shown are single exponential fits of data sets. (*Inset*) The midpoints 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 ≥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 ~69°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. ΔG° 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

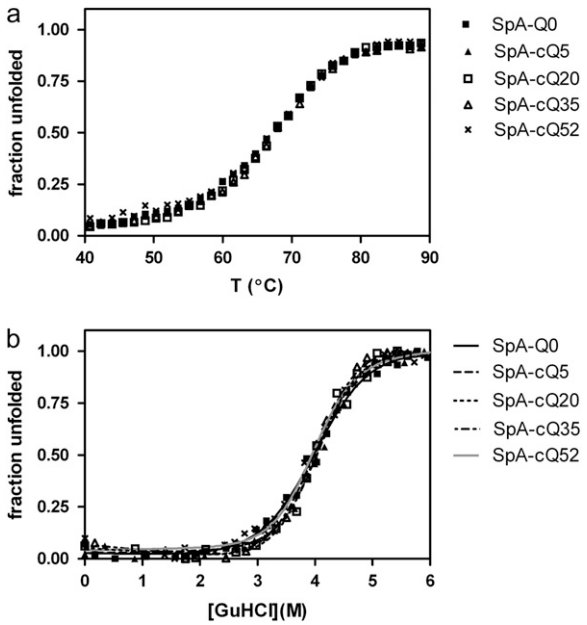

FIGURE 3 PolyQ expansion does not affect thermal or GuHCl-induced denaturation profiles. (*a*) All polyQ proteins showed similar thermal unfolding profiles. The change in CD signal at 222 nm was monitored as proteins were heated from 40°C to 90°C in the presence of 1 M GuHCl. (*b*) All polyQ proteins 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 overlayed (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 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 ^{1}H-^{15}N-HSQC experi-

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

	Tm (°C)*	Dm (M)†	$\Delta G°$ (kcal/mol)‡
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.
†Midpoints of the GuHCl-denaturation curves determined using the equation for two-state unfolding.
‡$\Delta G°$ was determined by analysis of the GuHCl-denaturation curves, with a shared equilibrium *m* value of 1.6 kcal/mol/M.

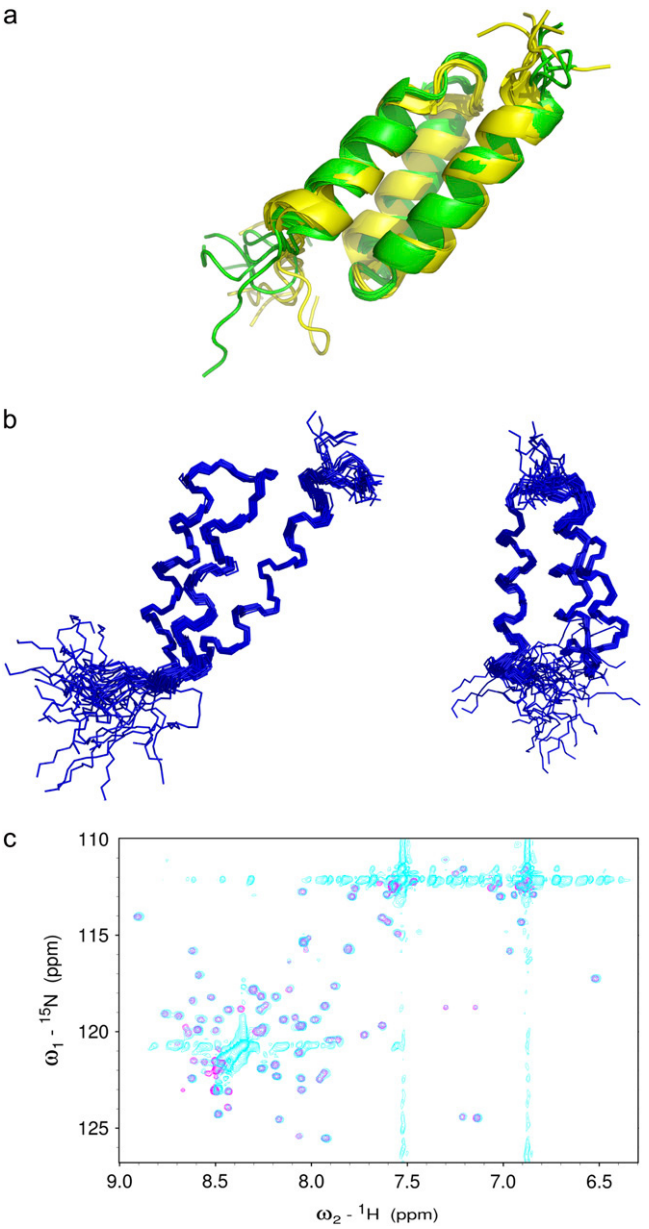

FIGURE 4 Structure of the flanking domain. (*a*) Backbone atom (N, Cα, C′) trace of the ensemble of 29 structures of SpA-cQ5 and 90° rotation around the *y* axis. (*b*) Backbone atom superposition of the five structures with the lowest energy from our ensemble of SpA-cQ5 (*green*) with the first five structures from the published 1SS1.pdb NMR ensemble for SpA (*yellow*). (*c*) Overlays of ^{1}H-^{15}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	77
ϕ	37
ψ	38
H-bond distance restraints	14
Structure statistics	
Violations (mean and SD)	
Distance constraints (Å)	0.0325 ± 0.0041
Dihedral angle constraints (°)	0.3187 ± 0.1571
Deviations from idealized geometry	
Bond lengths (Å)	0.0038 ± 0.00018
Bond angles (°)	0.7338 ± 0.0265
Impropers (°)	0.6903 ± 0.0354
Lennard-Jones energy (kcal mol^{-1})*	−258.91 ± 7.82
Root mean-square deviation to the mean structure (Å)	
Backbone	0.387 ± 0.124
Heavy	0.878 ± 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 ^{1}H and ^{15}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 ^{1}H-^{15}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 ^{1}H-^{15}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 α-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 ^{1}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^{17}, Leu^{18}, Leu^{20}, and Ala^{49}. The location of these methyl groups in the tertiary structure of SpA is highlighted in Fig. 6 *c*. The methyls

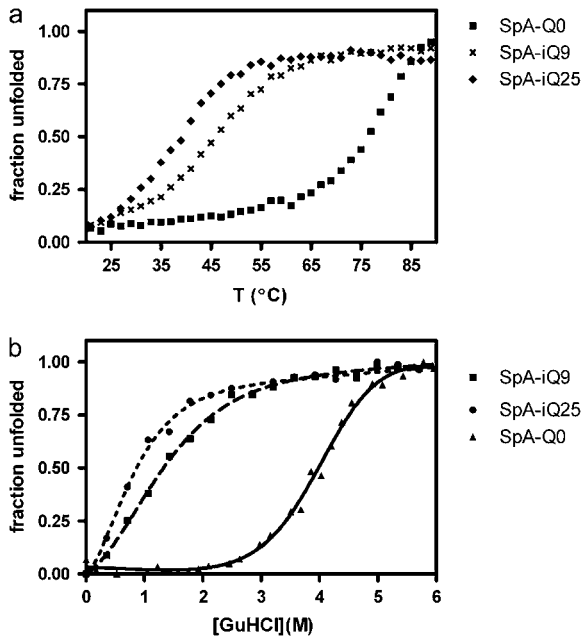

FIGURE 5 Intradomain polyQ tract affects the stability of SpA. (*a*) Thermal denaturation of SpA-Q0, SpA-iQ9, and SpA-iQ25. The change in CD signal was monitored as proteins were unfolded from 20°C to 90°C. (*b*) GuHCl unfolding of SpA-Q0, SpA-iQ9, and SpA-iQ25. SpA-iQ9 and SpA-iQ25 unfold in lower concentrations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

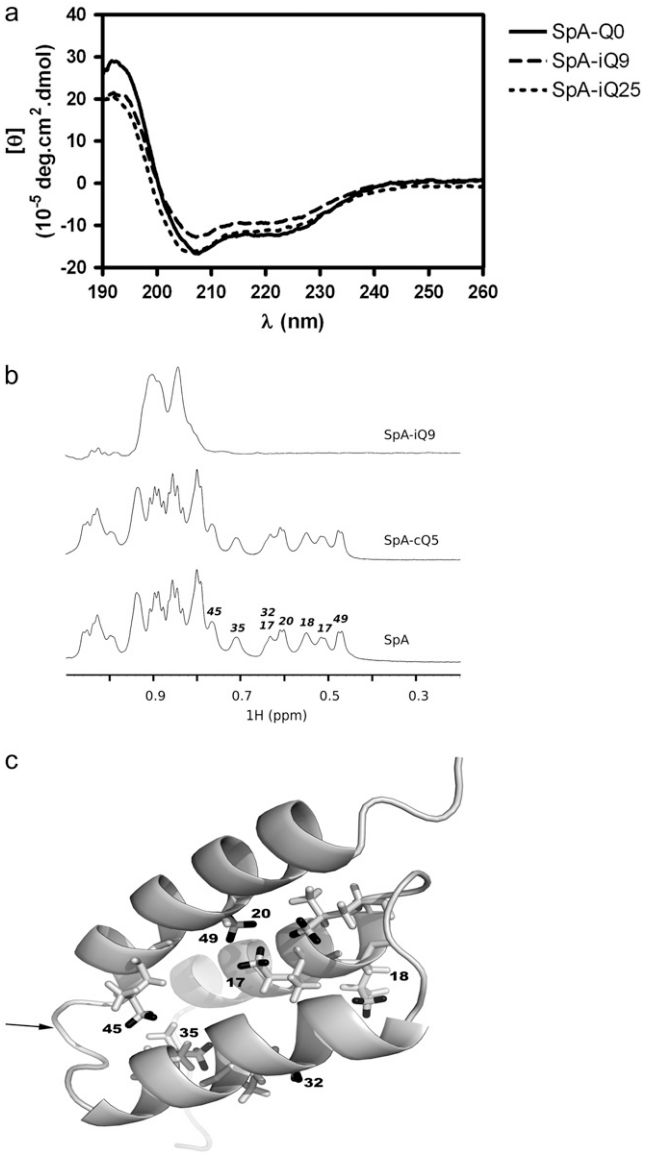

FIGURE 6 Intradomain polyQ repeat structurally perturbs SpA. (*a*) CD spectra of SpA-Q0, SpA-iQ9, and SpA-iQ25 showing that α-helical SpA secondary structure is retained in SpA-iQ9 and SpA-iQ25 variants. (*b*) The high-field region of the ^{1}H-NMR spectra of SpA-Q0, SpA-cQ5, and SpA-iQ9, respectively.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

of Ile^{17}, Leu^{20}, and Ala^{49} 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}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

	Tm (°C)*	Dm (M)†	$\Delta G°$ (kcal/mol)‡
SpA-Q0	—§	4.1	6.5 ± 0.2
SpA-iQ9	46.6 ± 1	1.5	1.8 ± 0.3
SpA-iQ25	38.9 ± 1	—¶	—¶

*Midpoints of thermal denaturation were determined by measuring the change in CD signal at 222 nm.
†Midpoints of the GuHCl-denaturation curves determined using the equation for two-state unfolding.
‡$\Delta G°$ was determined by analysis of the GuHCl-denaturation curves, with a shared equilibrium m value of 1.6 kcal/mol/M.
§Tm for SpA-Q0 could not be determined due to lack of post-transition baseline.
¶Dm and $\Delta G°$ 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 α-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 un/folding 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}H-^{15}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 β-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. Delanoy, 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 β barrel protein. *J. Biol. Chem.* 281:12959–12967.
20. Tanaka, M., I. Morishima, T. Akagi, T. Hashikawa, and N. Nukina. 2001. Intra- and intermolecular β-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 1H 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. Trottier. 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.