EXTENDED POLYGLUTAMINE TRACTS CAUSE AGGREGATION AND STRUCTURAL PERTURBATION OF AN ADJACENT β BARREL PROTEIN*

Zoya Ignatova‡ and Lila M. Gierasch

From the Departments of Biochemistry & Molecular Biology and Chemistry
University of Massachusetts-Amherst, Amherst, Massachusetts 01003

Running title: PolyQ Tracts Perturb Adjacent Domains

Address correspondence to: Lila M. Gierasch, Dept. of Biochemistry & Molecular Biology, Univ. of Massachusetts Amherst, Amherst, MA; Tel. 413-545-6094; Fax 413-545-1289; E-mail: gierasch@biochem.umass.edu

Formation of fibrillar intranuclear inclusions and related neuropathologies of the CAG-repeat disorders are linked to the expansion of a polyglutamine tract. Despite considerable effort, the etiology of these devastating diseases remains unclear. While polypeptides with glutamine tracts recapitulate many of the observed characteristics of the gene products with CAG repeats, such as in vitro and in vivo aggregation and toxicity in model organisms, extended polyglutamine segments have also been reported to structurally perturb proteins into which they are inserted. Additionally, the sequence context of a polyglutamine tract has recently been shown to modulate its propensity to aggregate. These findings raise the possibility that indirect influences of the repeat tract on adjacent protein domains are contributory to pathologies. Destabilization of an adjacent domain may lead to loss of function, as well as favoring non-native structures in the neighboring domain causing them to be prone to intermolecular association and consequent aggregation. To explore these phenomena, we have used chimeras of a well-studied globular protein and exon 1 of huntingtin. We find that expansion of the polyglutamine segment beyond the pathological threshold (> 35 glutamines) results in structural perturbation of the neighboring protein whether the huntingtin exon is N- or C-terminal. Elongation of the polyglutamine region also substantially increases the propensity of the chimera to aggregate, both in vitro and in vivo, and in vitro aggregation kinetics of a chimera with a 53 glutamine repeat follow a nucleation polymerization mechanism with a monomeric nucleus.

At least nine slowly progressing hereditary neurodegenerative diseases, including autosomal-dominant Huntington’s disease (HD1), dentatorubral and pallidoluysian atrophy, and several spinocerebellar ataxias, arise from the genetic expansion of an unstable polyglutamine (polyQ) repeat sequence.(1-3) Despite the functional and structural dissimilarity among the polyQ-containing proteins and the distinct pathologies they produce, each forms intranuclear inclusion bodies whose role in the disease pathogenesis is still controversial.(4) In all of the polyQ repeat diseases, there is a remarkable threshold effect of the length of the polyQ tract, which is inversely correlated with the age of onset and disease severity.(5) In the specific case of HD, the elongated CAG repeat encoding a stretch of glutamines is located within the N-terminal exon 1 of the gene encoding the 350-kDa protein ‘huntingtin’ ( htt). PolyQ repeats in the protein huntingtin (Htt) below the threshold value of 35 are not associated with HD; whereas repeats over 36 residues cause neurodegenerative dysfunction, and longer repeats are characterized by an earlier age of onset.(1) In ataxin-3, linked to type 3 spinocerebellar ataxia (Machado-Joseph disease), the polyQ repeat is located in the C-terminal portion of the protein, and a similar correlation of repeat length with disease severity and onset obtains.(6) The fact that disease risk and the tendency to aggregate in vitro show the same dependency on the number of glutamines supports a central role for polyglutamine-mediated aggregation in

Copyright 2006 by The American Society for Biochemistry and Molecular Biology, Inc.
disease pathogenesis. (1,7-9) Moreover, expression of htt exon 1 carrying the expanded CAG stretch recapitulates the toxicity in transgenic mice and C. elegans, causing an HD-like phenotype. (10,11) However, other studies argue that aggregate formation is not linked to disease pathology or cell death and may instead reflect cellular strategies to handle toxic species. (12-17) Fundamental cellular processes such as proteasome-mediated proteolysis have been proposed to be inhibited by polyQ-containing protein products with repeat lengths above the pathological threshold, (18-20) raising the possibility of pathological mechanisms apart from aggregation.

While much of the pathology of CAG repeat disorders can be attributed to the behavior of the polyQ segments, several studies have shown that the context of the polyQ tract must also be considered. (21-23) Recent work shows that the presence of a proline-rich extension C-terminal to a polyQ tract modulates its conformational behavior and decreases its tendency to aggregate. (24) In exon 1 of huntingtin, the glutamine-rich tract is followed by a proline-rich region, which based on this recent study, diminishes its aggregation propensity. In turn, effects on protein domains adjacent to glutamine-rich repeats may play a direct role in aggregation mechanisms. Incorporation of a Gly-Gln10-Gly peptide into chymotrypsin inhibitor 2 (CI2) led to formation of dimeric and trimeric oligomers by the normally monomeric CI2. (25) No perturbations in the structure of the CI2 were reported. By contrast, insertion of polyQ tracts longer than the pathological threshold into an otherwise stable protein (myoglobin) at the site of a surface-exposed loop caused modest structural perturbation and concomitant increase in aggregation propensity. (26,27) While in this case the structural perturbation was mild and did not appear to involve the core packing, the fact that myoglobin is very stable and contains a heme cofactor may have led it be more robust.

By comparison, Bevivino and Loll found that increasing the length of a glutamine repeat in the protein ataxin-3 from non-pathogenic (n=27) to a pathogenic length (n=78) was accompanied by significant structural destabilization as well as enhanced propensity to aggregate. (28) In order to carry out experiments on the protein carrying a 78 residue-long glutamine repeat, these authors used constructs with ataxin-3 linked to a solubility-enhancing fusion partner, maltose-binding protein. Their findings appeared to be supported by another study of a non-pathological variant of the protein ataxin-3, which was observed to form fibrillar aggregates upon partial denaturation using guanidine HCl. (29) This result implicates partially folded forms of protein domains neighboring polyQ tracts in the mechanism of aggregation. Recently, however, the same group has concluded that extending the polyQ tract in ataxin-3 does not destabilize the protein. (30) In the latter study the authors were forced to use pH denaturation and kinetics of unfolding/folding to obtain stability data as aggregation processes confounded the use of chaotropes such as guanidinium hydrochloride. They concluded that the length-dependent influence of the polyQ region was not on the native state of the flanking domain but rather on a fibrillogenic state. Moreover, they speculated that a flanking sequence may exert a restraining force on polyQ segments, modulating their aggregation propensities. Provocatively, in a recent study it is reported that the AXH domain of ataxin-3 has an intrinsic tendency to aggregate even in the absence of an adjacent polyQ tract. (23) All of these studies, while not uniform in their conclusions, point to the importance of understanding how polyQ tracts influence the sequences around them, and in turn how adjacent sequences may influence the behavior of polyQ repeats, in order to better understand CAG-repeat diseases.

Examining how a polyQ sequence may influence an otherwise stably folded adjacent domain has been limited by the technical difficulties of obtaining well-behaved soluble protein, in the presence of sequences that strongly favor aggregation. We have used fusions between cellular retinoic acid binding protein I (CRABP I) and Htt exon 1 for this work; constructs with the polyQ-containing sequence attached either at the N- or C-terminus of CRABP I were examined (Fig. 1A). The extensive previous characterization of this protein’s structure and folding (31-36) provides a foundation to interpret any changes caused by the polyQ fusions, and fortuitously, adequate solubility was obtained even with extended polyQ sequences to carry out biophysical study. CRABP I is a 136-amino acid long member of the intracellular lipid-binding protein (lLBP)
family. These proteins have a β barrel structure composed of 10-antiparallel β strands, with a short helix-turn-helix cavity between strand 1 and 2, wrapped around a ligand-binding cavity (solvent filled in the absence of ligand).\(^{(31,37)}\) In contrast to the recent study of ataxin-3\(^{(30)}\), we find that polyQ expansion within Htt exon 1 in the pathological range (repeat length of 53) causes structural perturbation and decreased stability of the neighboring CRABP I protein domain as well as increased propensity for aggregation both in vivo and in vitro, whereas a fusion of CRABP I with Htt exon 1 containing a shorter polyQ repeat (n=20) remains soluble with no evidence of structural disruption. The tight correlation of these effects with the polyQ length argues that the repeat tract is solely responsible for these perturbations and molecular pathologies. Aggregation in vitro of the CRABP I fusions to Htt exon 1 containing polyQ repeats above the pathological threshold showed nucleated polymerization kinetics and fit to a monomeric nucleus size, consistent with a dominant role for the polyQ sequence. We incorporated a tetra-Cys sequence within CRABP I in order to label the expressed protein in the cell with a low-molecular weight membrane-permeable fluorescent dye, FlAsH \((38)\). In vivo, this construct with pathological polyQ repeats (n = 40, 53 or 64) aggregated, and the apparent lag time could be modified by varying the amount of expressed protein, as expected for nucleated polymerization kinetics. This system provides a manipulable model to explore how polyglutamine tracts may influence juxtaposed folded domains; both conformational and energetic factors can be investigated due to the well-behaved nature of these constructs. We speculate that the nature of the CRABP I structure and its moderate stability may make it susceptible to perturbation from the polyQ sequences.

**MATERIALS AND METHODS**

**Constructs** - Tetra-Cys CRABP I (with a FlAsH-binding tetra-Cys motif) behind the T7 promoter was constructed as described.\(^{(39)}\) Upstream of the start codon comprising the NdeI restriction site, a new SalI restriction site necessary for the N-terminal insertion of the htt exon 1 was incorporated using a QuikChange protocol (Stratagene). For the C-terminal fusions, the stop codon and its nearest 3′-flanking region were replaced by Xhol and BamHI restriction sites. Due to the genetic manipulations, two new amino acids (Leu-Glu) were inserted at the C-terminus of tetra-Cys CRABP I in the C-terminal fusions. Sequences encoding htt exon 1 with CAG repeats of 20 and 53 were amplified by PCR with NdeI/SalI (lacking a stop codon) or Xhol/BamHI flanking recognition sites from plasmids pGEX-htt20Q and pGEX-htt53Q (gifts from U. Hartl and M. Hayer-Hartl, Max Planck Institute, Martinsried, Germany);\(^{(40)}\) sequences encoding htt exon 1 with 33, 40 and 64 length polyQ tracts were amplified with Xhol/BamHI flanking sites from plasmids pPD30.38-Q33-YFP, pPD30.38-40-YFP and pPD30.38-Q64-YFP, respectively (gifts of R. Morimoto, Northwestern Univ., Evanston, IL).\(^{(11)}\) The double-digested PCR products were cloned up- or down-stream of the tetra-Cys CRABP I gene yielding N- and C-terminal chimeras of tetra-Cys CRABP and Htt exon 1 with an N-terminal polyhistidine tag (sequence: MGHHHHHHHHHSSGHIEGRH-). The resulting plasmids were verified by DNA sequencing.

**Protein Expression and Purification** - All constructs (see Fig. 1A) were expressed in E. coli BL21(DE3) cells. Tetra-Cys CRABP I, Htt20 tetra-Cys CRABP (Htt exon 1 fused to the N-terminus), and tetra-Cys CRABP Htt20 (Htt exon 1 fused to the C-terminus) were purified from the soluble fraction of the cell lysates (39), whereas both the N- and C-terminal tetra-Cys CRABP fusions with Htt53 were isolated from the insoluble fraction after solubilization with 8M urea.\(^{(35)}\) Protein solutions (100-150 µM) of tetra-Cys CRABP I and tetra-Cys CRABP Htt20 (N or C) were stored at 4 °C and used within two weeks. Concentrations were determined spectrophotometrically using ε\(_{280}\) of 21750 M\(^{-1}\)cm\(^{-1}\) (a corrected value for wild-type CRABP I containing four additional cysteines). Refolding was accomplished by drop-wise dilution (100-fold) of the urea-containing protein fractions in 10 mM Tris.HCl, pH 8.0 or in 10 mM HEPES, pH 7.8 containing 2 mM β-mercaptoethanol (depending on the subsequent step). The refolding efficacy was determined by ligand (retinoic acid) binding, CD-spectroscopy and measurement of the intrinsic Trp fluorescence. The variants with 53 polyQ
showed a high tendency to aggregate and were stored in dilute stocks (10-15 nM). The protein samples were proteolytically stable for at least 6 days, as evidenced by mass spectrometry. Before use small aliquots were concentrated to the desired concentration using a Centricon system (molecular weight cut-off 10 kDa) at 4 °C and centrifuged at 14000 rpm in a microfuge for 10 min to remove aggregates. The properties of refolded tetra-Cys CRABP I Htt53 (N- or C-terminal fusion) were found to be the same as those of protein isolated in very low yield from the soluble fraction 45 minutes post-induction.

**In-cell FLAsH-EDT2 Labeling - E. coli BL21(DE3)** cells expressing tetra-Cys constructs were labeled as described.(39) A 10-fold excess of ethane dithiol (EDT) was added to suppress the labeling of the endogenous cysteine pairs.(38) The cells were preloaded with FLAsH-EDT2 one generation prior to induction with 0.4 mM IPTG at A 600=1. Aliquots of 150 µl were withdrawn and subjected to fluorescence measurements at 530 nm (bandwidth 2 nm) with excitation at 500 nm (bandwidth 2 nm) (Photon Technology International QM-1 fluorometer). The fluorescence of a negative control of CRABP I wild-type expressing cells pre-loaded with FLAsH-EDT2 was subtracted from each point. [N.B. FLAsH is now marketed by Invitrogen under the trade name ‘Lumio’.] Cell Fractionation - 10 ml aliquots of cells expressing the tetra-Cys CRABP I variants were harvested by centrifugation and fractionated into soluble and insoluble fractions.(39) The insoluble fraction was solubilized in 50 mM phosphate buffer pH 8.0 containing 8M urea and 2% SDS and centrifuged at 27,000 x g for 30 min. The pellet was re-suspended in 1.5 ml of 20 mM Tris.HCl buffer and defined as the SDS-insoluble fraction. All fractions were analyzed on a 12% SDS-PAGE, and the intensity of the Coomassie-stained bands of interest was detected by optical densitometry (Bio-Rad).

**Circular Dichroism and Fluorescence -** Far-UV CD spectra were collected at 20 °C on a Jasco J-715 spectropolarimeter with a 1 mm path length cell (1 nm bandwidth). The concentration of protein was 5 µM, and each spectrum was an average of five spectra. Fluorescence spectra of 3 µM proteins were taken at 20 °C with an excitation at 280 nm (1 nm slit width) and emission from 300-380 nm (2 nm slit width).

**Stability Measurements -** For the urea unfolding titrations, 3 µM protein (FIAsH-labeled) was diluted in 0 to 8M urea containing 10 mM Tris.HCl pH 8.0 and incubated for >2 hours at 37 °C. Unfolding was monitored at 37 °C by Trp fluorescence (as above). Urea titration curves were fit to a two-state model to derive thermodynamic values. To test proteolytic susceptibility and ability to bind ligand, tetra-Cys CRABP I variants were digested with α-chymotrypsin (α-CT:protein=1:20 ratio) for 45 minutes on ice, with and without pre-incubation with 2.5-fold excess of retinoic acid (RA) (5 min in the dark). The reaction was stopped by adding SDS-loading buffer and boiling.

**Aggregation Time-course Assays In Vitro -** Aliquots of tetra-Cys CRABP I Htt53 (10-15 nM) were labeled with FIAsH-EDT2 for 60 min as described previously.(41) The diluted labeled tetra-Cys CRABP I Htt53 solutions were concentrated at 4 °C by ultrafiltration (cut-off 10 kDa) and centrifuged at 14000 rpm in a microfuge for 10 minutes at 4 °C. Solutions were used immediately to conduct kinetic experiments. In order to study the aggregation process in the laboratory at reasonable time scales, the labeled protein was destabilized with 1 M urea (in 10 mM HEPES pH 7.8). Aggregation was followed by increasing FIAsH fluorescence or by estimation of the concentration of soluble monomer by absorbance measurements of the clarified supernatant at 280 nm as described.(41) In both fluorescence and spectroscopic measurements, the samples were incubated without stirring and gently vortexed prior to withdrawal of an aliquot to insure an equal distribution of the aggregates over the entire volume of the sample. In parallel, the fluorescence and absorbance at 280 nm of controls without protein were recorded for all experiments and subtracted from those of the samples.

**Seeding Experiments -** 10 µM FIAsH-labeled tetra-Cys CRABP Htt53 was destabilized by dissolving in 1.0 M urea and incubated overnight at 37 °C without stirring to allow complete aggregation. To improve their seeding potency, the aggregates were sonicated for 2 min in 30 s bursts after washing with equal volumes of water and 10 mM Tris.HCl pH 8.0. Solutions of labeled tetra-Cys CRABP I Htt53 were pre-incubated to reach the test temperature of 37 °C, then seeded with...
pre-formed aggregates representing 1%, 5% and 10% of the initial tetra-Cys CRABP I Htt53 concentration. Aggregation was monitored by FIAsH fluorescence.

**Determination of Nucleus Size** - The time-course of *in vitro* aggregation was analyzed by the mathematical model describing nucleation-controlled aggregation kinetics developed by Ferrone (42) that has been applied to several cases of protein aggregation (7,41,43). A plot of the initial rate of aggregate formation (followed by FIAsH fluorescence) or the changes of the soluble monomer concentration vs. $t^2$ gives a slope equal to $1/2 k_+^2 K_{n^*} c^{n^*+2}$, where $k_+$ is the elongation rate, $K_{n^*}$ is the equilibrium constant for the monomer-nucleus equilibrium, $c$ is the concentration of monomer, and $n^*$ is the number of monomers in the nucleus. Plotting the log of the slope versus the log of the bulk monomer concentration \[ \ln(\text{slope}) = \ln\left(\frac{1}{2} k_+^2 K_{n^*}\right) + (n^*+2) \ln(c) \] yields a new slope of $n^*+2$, which can be used to estimate the size of the critical nucleus.

**Fluorescence and Electron Microscopy Experiments** – Tetra-Cys CRABP I Htt53 expressing cells labeled with FIAsH-EDT$_2$ were harvested at different times after induction and imaged by fluorescence microscopy. Intact inclusion bodies were isolated from tetra-Cys CRABP I Htt53 expressing cells labeled with FIAsH-EDT$_2$ and harvested four hours post-induction, as described by Oberg et al.(44) For fluorescence microscopy, two µl of labeled cells or aggregates isolated from cells were immobilized in 1% agarose in LB and imaged using a Nikon Eclipse E600 with excitation at 485 nm and a 510 nm emission cut-off filter. For electron microscopy, samples of isolated aggregates were applied to carbon coated copper grids, stained with 3% (w/v) ammonium molybdate (pH 6.8) and imaged under defocus conditions with a Phillips CM12 operating at 120 kV.

**RESULTS**

**Htt exon 1 with an expanded polyQ tract causes aggregation *in vivo***. Expression of the fusion proteins in which Htt exon 1 with varying lengths of polyQ tracts was attached to the C-terminus of tetra-Cys CRABP I (Fig. 1A) was followed by fluorescence at 530 nm (excitation 500 nm) after pre-loading the cells with FIAsH dye as described previously.(39) The results show that these constructs exhibit polyQ-length dependent aggregation propensities. The fluorescence of cells expressing labeled tetra-Cys CRABP I Htt exon 1 fusions as a function of time after induction is close to that of cells expressing the soluble tetra-Cys CRABP I in the first 200 minutes when the polyQ repeat is in the non-pathological range (20 and 33) (Fig. 1B). By contrast, a markedly different time course is observed when the polyQ repeat length exceeds the pathological threshold, *i.e.*, when $n=53$ and 64 (Fig. 1B). In these two cases, there is essentially no lag from the time of induction to the time at which a rapid increase in fluorescence occurs; we have demonstrated in previous work that the abrupt fluorescence increase correlates with formation of inclusion bodies.(39,41) We confirmed that the steep fluorescence increase reports on aggregation by carrying out cell fractionation (Fig. 1C). The tetra-Cys CRABP Htt20 variant remained soluble during the entire expression cycle and was present in amounts comparable to the amounts of tetra-Cys CRABP I, based on fractionation results (Fig. 1C); a small amount of aggregate was observed for the construct with 33 glutamines. While the tetra-Cys CRABP Htt40 still partitions between the soluble and insoluble fractions, both tetra-Cys CRABP I fusions with longer polyQ stretches (53 and 64) were found almost exclusively in the insoluble fraction. A noteworthy observation is that for the shorter polyQ repeats ($n=33$ or 40) all the aggregated protein was SDS soluble, whereas the aggregates of the chimeras with 53- and 64-residue long polyQ repeats partitioned between SDS-soluble and SDS-insoluble fractions (Fig. 1C). The longer the polyQ stretch the higher the amount of the SDS-insoluble aggregates. The property of insolubility in SDS is associated with the formation of fibrillar aggregates.

We next asked if the position of the polyQ-containing Htt exon 1 in the chimera altered its influence on the fusion protein’s aggregation propensity. The fluorescence time course for N-terminally positioned Htt 20 tetra-Cys CRABP I and Htt 53 tetra-Cys CRABP I expression were essentially the same as when the
corresponding Htt exon 1 segments were added C-terminally (Fig. 2A,B).

The absence of a lag-phase in the \textit{in vivo} kinetic curves for the Htt 53 and 64 fusion proteins (within the time resolution of the experiment) suggests that aggregation is favorable essentially as soon as expression of protein commences. Reducing the rate of translation and thereby lowering the protein concentration by treatment with low concentrations of chloramphenicol (2.5 µg/ml) led to a short lag-phase whose duration increased at higher chloramphenicol concentration (3.5 µg/ml) (Fig. 2C). Under these conditions the \textit{in vivo} aggregation curve took on the typical sigmoidal shape characteristic of amyloidogenesis (45) and mirrored \textit{in vitro} results (see below).

The aggregates formed by the Htt exon 1 chimeras with longer polyQ repeats \textit{in vivo} changed character over time. As shown in Figure 2B for the C-terminal fusion, within the first 30 min, a small amount of the protein aggregated, and maximal insoluble material was found at 120 min. At this time point, fluorescence microscopy showed hyperfluorescent aggregates at the poles of the bacterial cells (Fig. 2D). At later time points (>150 min after induction), the isolated insoluble aggregates underwent an interesting metamorphosis and converted into detergent-resistant aggregates, detectable between the stacking and resolving gels on SDS PAGE. Aggregates isolated before this conversion appeared amorphous with spherical ensembles in electron microscopy (data not shown). Fluorescence and electron microscopic inspection of the detergent-resistant aggregates isolated from cells suggested, by contrast, that they are fibrillar (Fig. 2E, F). The transition in the nature of the aggregates paralleled an increase of the fraction of cells with a filamentous phenotype, which is associated with a stress response and is reminiscent of the reported toxicity of Htt fragments with expanded polyglutamine segments in eucaryotic organisms.(11,46)

\textit{In vitro} aggregation of tetra-Cys CRABP I Htt53 follows a nucleation-polymerization mechanism through a monomeric nucleus. Although not amenable to a detailed quantitative analysis, the \textit{in vivo} kinetic data described above are consistent with a nucleation polymerization mechanism with an apparently very short nucleation step when the polyQ tract is longer than 40. The high synthesis rates in bacteria in the expression system used appear to have masked the initial nucleation step, since some lag time is observed when synthesis is inhibited (Fig. 2D), suggesting a very low threshold concentration for aggregate propagation. To test this hypothesis, to carry out quantitative analysis of aggregation kinetics, and to further explore the aggregation mechanism, we examined the aggregation of tetra-Cys CRABP Htt53 \textit{in vitro}. Aggregation kinetics were followed by two different methods: FIASH-fluorescence and changes in the concentration of the soluble monomer. Based on our previous work (41), we correlated the observed hyperfluorescence with the production of non-native states including aggregates. The FIASH signal is dominated by the aggregates, and thus FIASH fluorescence can be used to follow aggregation \textit{in vitro}. The use of FIASH fluorescence to quantitate aggregate formation is supported by comparing results based on measurement of soluble monomer concentration. Aggregation of the tetra-Cys CRABP I Htt53 chimera under native conditions at 37 °C is a relatively slow process. To accelerate aggregation kinetics to an extent that facilitated measurements at a reasonable time scale, we chose to work at an intermediate urea concentration (1M), where partially folded states are populated significantly. The choice of conditions was based on urea-induced unfolding of tetra-Cys CRABP I Htt53 monitored by intrinsic Trp fluorescence, which showed an apparent denaturation midpoint of 1.05±0.08 M (see below and Table 1). The observation of faster aggregation near the denaturation mid-point has been made for several other proteins and exploited in protocols for following aggregation kinetics.(47) Note that this observation points to a likely mechanism of aggregation with an active participation by CRABP I, since the urea concentration is not expected to have as significant an effect on the polyQ tract as it does on the CRABP I. Under the same conditions, the fusion between tetra-Cys CRABP I and Htt exon 1 in the non-pathological range (n=20) did not aggregate (Fig. 3A).

\textit{In vitro} aggregation of tetra-Cys CRABP I Htt53 showed a sigmoidal time-dependence, with a well defined lag-phase followed by an exponential increase of the FIASH-fluorescence (Fig. 3A) and decrease of the soluble monomer
(Fig. 3B). These observables plateaued at later time points. Increasing the initial concentration of the bulk monomer shortened the lag-phase, and the exponential growth reached higher final values of the fluorescence in the plateau phase (Fig. 3A). Overlaying the corresponding curves from both experiments shows a good inverse correlation, and the exponential increase of FlAsH fluorescence, reporting on the aggregate growth, coincides with the monomer disappearance.

More conclusive evidence for nucleation-polymerization aggregation was provided by quantitative analysis of kinetics during the initial time period. The time course of aggregate formation followed by FlAsH-fluorescence or disappearance of the soluble monomer could be fitted using methods developed by Ferrone (42) for nucleated polymerization, since the initial aggregation phase was observed to be a function of time square ($t^2$). Plotting the log of the aggregation rate versus the log of the monomer concentration gives a quantitative measure of the nucleus size with $n*$ monomeric units (Fig. 3C; see Materials and Methods and references 7 and 41). Applying this analysis to the initial aggregation rates deduced from the aggregation kinetics followed by the FlAsH fluorescence (Fig. 3A) and by the decrease of the monomer concentrations (Fig. 3B) at different concentrations of tetra-Cys CRABP I Htt53 yielded slopes ($n^* + 2$; where $n^*$ is the number of monomers in the nucleus) of 2.86 and 2.90, respectively, consistent with a monomeric nucleus. Interestingly, a monomeric nucleus was previously implicated in aggregation of a mutant form of CRABP I (41) and in peptides with glutamine repeats (7), making it difficult to distinguish which species is mechanistically dominant.

Another line of evidence supporting a nucleation-dependent polymerization mechanism is the effect of pre-formed seeds on the aggregation rate: Low seed concentrations (1% (w/w)) shortened the apparent lag-phase significantly, whereas seeds at 2.5% and 5% (w/w) were able to completely abrogate the well-defined lag-phase (Fig. 3D).

Linear regression of the final plateau level of soluble monomer concentration for varying initial monomer concentrations and extrapolation of the $x$-intercept provides an estimate of the critical concentration for aggregation.(45,48) For all initial tetra-Cys CRABP I Htt53 concentrations (1-20 µM), the amount of the soluble monomer decreases progressively as a function of time reaching at late times an equilibrium concentration of 33-50 nM (Fig. 3E), which provides an estimate of the critical concentration for aggregation of this construct.

### Presence of an N- or C-terminal adjacent expanded polyQ tract destabilizes and perturbs the structure of tetra-Cys CRABP I.

The spectral features of the chimeras between tetra-Cys CRABP and Htt20 are nearly identical to those of tetra-Cys CRABP I (see Fig. 4A-D). Strikingly, expansion of the polyQ tract from 20 to 53 residues perturbs the CRABP tertiary and secondary structure, whether the Htt exon 1 domain is added to the N- or C-terminus (Fig. 4A-D). For example, the fluorescence of intrinsic Trp residues in native tetra-Cys CRABP I Htt53 was shifted from 328 nm (characteristic of stably folded tetra-Cys CRABP I) to 336 nm, indicating significantly more solvent exposure and changes in the local Trp environments (Fig. 4A). These spectral properties are reminiscent of partially unfolded species of tetra-Cys CRABP I induced by intermediate concentration of denaturant. The specific inflection at 228 nm in the tetra-Cys CRABP I CD spectrum, previously shown to arise from the fine tertiary structure surrounding Trp87 and Trp109,(49) disappeared in the CD spectrum of the tetra-Cys CRABP I fusions to Htt53, and the minimum of the spectrum shifted to shorter wavelengths (Fig. 4C, D), also suggesting partial unfolding, a gain of random coil structure, and an increased conformational heterogeneity (particularly given the broader minimum). Addition of the native ligand retinoic acid stabilized the structure of both N- and C-terminal fusions of tetra-Cys CRABP I with Htt53, as indicated by a shift of the minimum of the CD spectrum to 218 nm, characteristic of the native β-sheet structure, and a decrease of the fluorescence intensity accompanied by a blue-shift towards the native maximum (to 332 nm). Addition of retinoic acid protects CRABP I against chymotryptic digestion, and this protection is partially maintained in tetra-Cys CRABP I Htt53 (Fig. 4E). All of these ligand-dependent effects argue that the fusion proteins with the extended polyQ tract retain some capacity to bind ligand, and that the long polyQ tract reversibly perturbs the fold of its...
neighboring CRABP domain. The added stability from ligand binding counteracts the destabilizing effect of the Htt exon 1 with an extended polyQ tract.

Urea-induced equilibrium unfolding titrations of tetra-Cys CRABP I fusions to Htt20, either N- or C-terminal, are very similar to that for tetra-Cys CRABP I (Fig. 5), with comparable free energies between native and urea-unfolded states (Table 1). Expansion of the polyQ tract to 53 residues led to significantly lower apparent stability in urea melts. Quantitative comparison of free energies was complicated by the pronounced tendency of the tetra-Cys CRABP I/Htt53 fusion proteins to aggregate. At higher urea concentrations, melts showed deviations from the sigmoidal shape expected for reversible denaturation. Aggregation was less of a problem at the lower urea concentration branches of the denaturation curves, enabling us to derive apparent half-maximal urea concentrations for both the N- and C-terminal Htt53 fusions; these c_m’s were lower than those for either the tetra-Cys parent protein or the Htt20 variants by nearly 1 M (Table 1), arguing for substantially lower stability. By assuming reversible denaturation, we could estimate that the extended polyQ tract led to 1.5 kcal/mol loss of equilibrium stability (Table 1).

DISCUSSION

Proteins involved in polyQ disorders share little or no structural or functional homology except for the presence of a polyQ region somewhere in their sequences. Consequently, the similar neuropathological picture in the various CAG repeat disease states has generally been ascribed to the presence of an expanded polyQ tract.(2,3) The panoply of effects of the polyQ expansions on the host protein and its interactions has been difficult to assess. A dominant role of the polyQ tract in favoring aggregation is well established in model systems,(7) and the ability of polyQ peptides to lead to symptoms akin to the pathological changes in CAG-repeat diseases has been shown in mice (10) and C. elegans (11). However, several possible mechanistic schemes remain under consideration in the etiology of the neurodegenerative diseases. Alterations of the functions and interactions of proteins hosting polyQ tracts must be included along with direct toxic effects of aggregates as possible contributors to disease. Also, the aggregates that form in cellular conditions may include all or part of the host protein (due to partial proteolysis) and may be formed both because of intrinsic properties of polyQ segments(22) and because of partial unfolding of host proteins, which then become more aggregation prone. In order to gain a better understanding of these alternatives, fundamental biochemical and biophysical data are needed on the impact of extended polyQ tracts on host proteins, and reciprocally, how host sequences may modulate the behavior of the polyQ sequence.

This study was directed towards the goal of assessing how a polyQ repeat affects a neighboring otherwise stably folded globular protein. Our data using the β clam protein, CRABP I, as a model host protein clearly show that the lengthening of the polyQ tract in the Htt exon 1 beyond the pathophysiological threshold led to marked propensity of the fusion proteins to aggregate and a substantial perturbation of the structural integrity of the adjacent model protein. We found the fusions of tetra-Cys CRABP I and Htt exon 1 with non-pathological length of the polyQ tract (20 or 33 consecutive glutamines) to be completely soluble when expressed in E. coli, whereas the tetra-Cys CRABP I fusions with Htt exon 1 containing a stretch of 40, 53, or 64 polyglutamines were highly aggregation-prone. Moreover, the fusion of an Htt exon 1 with an expanded polyQ tract (n=53) to tetra-Cys CRABP I perturbs and destabilizes the native conformation of the neighboring globular protein domain. The Htt53-induced perturbation of the spectroscopic properties of tetra-Cys CRABP suggests a partial unfolding, leading either to a homogeneous population of imperfectly folded species or to a heterogeneous population with some species more native-like and some partially unfolded. Here, the simple fact that a polyQ-containing exon with a high repeat number is neighboring to the CRABP disrupts its native-state structure, which likely contributes to enhanced propensity to aggregate. While the destabilization caused by adjacency to the Htt exon 1 may favor aggregation, we cannot distinguish aggregation mechanisms mediated primarily by the partially folded CRABP I component or by the polyQ component, since studies on both of these show nucleation-polymerization kinetics with a monomeric
We are following up on this question by further characterization of the nature of the aggregated species formed by the chimeras. Intriguingly, the observed change in the character of the aggregates formed in vivo by the chimera with an expanded polyQ tract from detergent soluble to detergent insoluble, accompanied by a morphological change to fibrillar character appears to parallel a change from CRABP I-dominated aggregation to polyQ-dominated aggregation. The preliminary evidence for this is the observation that the earlier aggregates can seed CRABP I aggregation but not polyQ, and the later aggregates can seed polyQ aggregation, but not CRABP I (data not shown). It will be exciting to explore in greater depth how the observed structural perturbation and the mechanism of aggregation are related to one another.

What is the physical basis for the destabilization of a globular protein by polyQ sequences? Two formal possibilities must be considered: 1) the tendency of the polyQ segment to form intermolecular associations leads to a conformational constraint on adjacent sequences because of the driving force to draw segments into antiparallel β sheet, for example as reported for myoglobin with a 50-glutamine insertion; or 2) the intramolecular conformational propensities of the polyQ segment compete with the adjacent sequence’s ability to participate in native interactions critical to the stably folded protein. This could be either by the preference of the polyQ segment for a particular structure, or by indirect effects it has, for example, by stabilizing the denatured state of its neighbor. The work done early on by Perutz and coworkers with oligo-Gln insertions in CI2 argues that short stretches of glutamine can be tolerated without perturbing adjacent sequence, although they saw increased tendency for oligomers. (25) While one might have predicted that these oligomers would be stabilized by polyQ/polyQ interactions, as in the case of the polyQ-inserted myoglobin, instead a later crystal structure showed that domain swapping between the CI2 monomers mediated dimer formation.(50) This form of interaction suggests that there was a destabilizing influence of the oligo-Gln segment on the CI2, similar to our findings. Analogously, structural perturbation and increased aggregation of ataxin-3 were reported by Bevivino and Loll when the adjacent polyQ segment was extended from 27 to 78.(28) This latter study was hampered by insolubility of the protein with the longer glutamine repeat, necessitating use of a solubility-enhancing fusion protein, and now has been called into question by the recent work using acid denaturation and unfolding/refolding to assess stability changes.(30) This latter study also was challenged by the insolubility of the constructs with extended polyQ repeats.

An advantage of the system we designed is that adequate solubility was retained in the fusions with extended polyQ repeats to enable measurements of some biophysically useful parameters, and possibly even, in the long run, structural elucidation. A similar fusion construct in which extended polyQ segments were soluble enough for biophysical characterization was reported by Masino et al. (51) and enabled the random coil character of the polyQ to be deduced. In our construct, we can monitor the polyQ segment and the neighboring domain simultaneously. The measurements we have made to date allow us to say something about how destabilizing the polyQ-containing tract is. The estimated change in stability of tetra-Cys CRABP I upon fusion to Htt exon 1 with 53 glutamines was 1.5 kcal/mol at 37 °C. We hypothesize that two factors contribute to the impact of the flanking extended polyQ-containing Htt exon 1 sequence on CRABP I: the modest native state stability of CRABP I at that temperature (3.6 kcal/mol), which is insufficient to offset the conformational energetic driving force of the polyglutamine stretches longer than 20; and the potential susceptibility of the β barrel structure of iLBPs to disruption by perturbation of terminal sequence elements.(52) Hence, an overall structural perturbation (loss of native structure) was observed. A major limitation in our understanding is the lack of structural knowledge about polypeptides composed of glutamines. Several models for the preferred conformation of polyQ-containing proteins have been proposed.(53-58) There is general agreement that short oligoglutamine stretches are unstructured, but that β structure is strongly preferred by longer polyQ tracts. The β structure has been proposed to exist in β hairpins, β sheets (parallel or anti-parallel), or in β helices. In general, these structures can form
intramolecularly or intermolecularly, and the balance between these modes of formation may be affected by the length of the glutamine repeat. However, estimates of the energetic driving force towards formation of these specific structures are not available. Assuming additivity and a linear dependence on glutamine number, our data suggest that each glutamine after 20 might contribute 0.05 kcal/mol of conformational free energy that can bias the adjacent chain’s behavior. Further experiments are clearly needed to pin down these effects quantitatively.

An analogy for the thermodynamic ‘tug-of-war’ between CRABP I and the polyQ-containing Htt exon 1 suggested by our results is the behavior of an engineered chimera between ubiquitin and barnase (59): Because of the design of the insertion of ubiquitin into barnase, the topology of their native states are incompatible with retention of native structure by both components of the chimera. Loh and coworkers showed that the higher free energy of stability for barnase upon binding barstar unfolded the ubiquitin, but in the absence of barstar, ubiquitin remained folded and barnase lost its native structure. Because of its barrel topology, CRABP I may be exquisitely sensitive to truncation at either its termini. It is known that a three-residue truncation of a related iLBP leads to destabilization of the native state. (52) The terminal sequences in this protein are close and form several mutual interactions. Thus, the conformational tendency of Htt exon 1 with an extended polyQ segment could be sufficient to disrupt key interactions in the terminal regions of the CRABP I β barrel. In future work, it will be crucial to determine the structure of the polyQ segment and the mechanism by which it disrupts the CRABP I native state. Its effect indeed is likely to be intramolecular, since our experimental protocols were optimized to favor monomeric products.

The idea that polyQ tracts would favor a conformation strongly enough to perturb an adjacent protein’s structure, without forming intermolecular interactions, seems at variance with thermodynamic estimates of the energetics of formation of aggregation nuclei from Wetzel and coworkers (60). Their studies suggest that adoption of the conformation that nucleates aggregation is a very uphill thermodynamic process with a very small equilibrium constant. To reconcile these results with the model that the polyQ tract can affect an adjacent domain in an intramolecular fashion would require that the conformation adopted by the polyQ sequence as it perturbs the flanking protein is not the one that nucleates aggregation.

A complete understanding of the process of disease-related amyloid formation relies on information about the mechanism of aggregation directly in the cell. The approach we describe here has enabled us to monitor aggregation in real time directly in cells and in so doing to observe behaviors that depend on the length of the homopolymeric polyQ region comparably to the pathologies described for CAG repeat diseases. It is of interest to compare the processes we are able to characterize in vitro with those that occur under the solution conditions that obtain in the cell in order to relate the time courses of aggregation. We estimated from our in vitro data that the critical concentration below which no aggregation of tetra-Cys CRABP I Htt53 occurs is between 33 and 50 nM. The total concentration of tetra-Cys CRABP Htt53 in E. coli cells 30 minutes after induction can be estimated to be 1.75 µM using optical densitometry and fractionation results (Fig. 2B). Assuming a constant synthesis rate, the amount of the newly synthesized tetra-Cys CRABP I Htt53 in E. coli cells would rapidly approach and exceed the critical concentration for aggregation in vitro. Moreover, the crowded environment in the cell is expected to reduce the critical concentration. (24) It is therefore not surprising that we observe aggregation as soon as we monitor samples after induction of the protein and that we therefore see no apparent lag-phase.

The system we describe sheds light on the marked impact of a naturally-occurring polyQ-containing sequence, exon 1 of huntingtin, on a model protein that is adjacent to it, either upstream or downstream. The labeling strategy we have used enables the observation of all of these events and behaviors in vivo, with the added benefit that it is possible to screen for the effects of environment and added small molecules on these events. Because this system makes the direct observation of aggregation events possible, it offers advantages over some recently described screens for aggregation inhibitors that are less direct. (61, 62)
REFERENCES

1. Zoghbi, H. Y., and Orr, H. T. (2000) Annu Rev Neurosci 23, 217-247
2. Ross, C. A., and Poirier, M. A. (2004) Nat Med 10 Suppl, S10-17
3. Cummings, C. J., and Zoghbi, H. Y. (2000) Annu Rev Genomics Hum Genet 1, 281-328
4. Ross, C. A. (2002) Neuron 35, 819-822
5. Margolis, R. L., and Ross, C. A. (2001) Trends Mol Med 7, 479-482
6. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., et al. (1994) Nat Genet 8, 221-228
7. Chen, S., Ferrone, F. A., and Wetzel, R. (2002) Proc Natl Acad Sci U S A 99, 11884-11889
8. Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G. P., Lehrach, H., and Wanker, E. E. (1999) Proc Natl Acad Sci U S A 96, 4604-4609
9. Trottier, Y., Lutz, Y., Stevanin, G., Imbert, G., Devys, D., Cancel, G., Saudou, F., Weber, C., David, G., Tora, L., and et al. (1995) Nature 378, 403-406
10. Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S. W., and Bates, G. P. (1996) Cell 87, 493-506
11. Morley, J. F., Brignull, H. R., Weyers, J. J., and Morimoto, R. I. (2002) Proc Natl Acad Sci U S A 99, 10417-10422
12. Simeoni, S., Mancini, M. A., Stenoien, D. L., Marcelli, M., Weigel, N. L., Zanisi, M., Martini, L., and Poletti, A. (2000) Hum Mol Genet 9, 133-144
13. Nucifora, F. C., Jr., Ellerby, L. M., Wellington, C. L., Wood, J. D., Herring, W. J., Sawa, A., Hayden, M. R., Dawson, V. L., Dawson, T. M., and Ross, C. A. (2003) J Biol Chem 278, 13047-13055
14. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R., and Finkbeiner, S. (2004) Nature 431, 805-810
15. Saudou, F., Finkbeiner, S., Devys, D., and Greenberg, M. E. (1998) Cell 95, 55-66
16. Hodgson, J. G., Agopyan, N., Gutekunst, C. A., Leavitt, B. R., LePiane, F., Singaraja, R., Smith, D. J., Bissada, N., McCutcheon, K., Nasir, J., Jamot, L., Li, X. J., Stevens, M. E., Rosemond, E., Roder, J. C., Phillips, A. G., Rubin, E. M., Hersch, S. M., and Hayden, M. R. (1999) Neuron 23, 181-192
17. Gutekunst, C. A., Li, S. H., Yi, H., Mulroy, J. S., Kuemmerle, S., Jones, R., Rye, D., Ferrante, R. J., Hersch, S. M., and Li, X. J. (1999) J Neurosci 19, 2522-2534
18. Venkatraman, P., Wetzel, R., Tanaka, M., Nukina, N., and Goldberg, A. L. (2004) Mol Cell 14, 95-104
19. Park, Y., Hong, S., Kim, S. J., and Kang, S. (2005) Mol Cells 19, 23-30
20. Holmberg, C. I., Stanisiewski, K. E., Mensah, K. N., Matouschek, A., and Morimoto, R. I. (2004) Embo J 23, 4307-4318
21. Nozaki, K., Onodera, O., Takano, H., and Tsuji, S. (2001) Neuroreport 12, 3357-3364
22. Wetzel, R. B. (2006) in Genetic Instailities and Neurological Disorders (Wells, R., and Ashizawa, T., eds), 2nd Ed., Elsevier
23. de Chiara, C., Menon, R. P., Dal Piaz, F., Calder, L., and Pastore, A. (2005) J Mol Biol 354, 883-893
24. Bhattacharyya, A., Thakur, A. K., Chellgren, V. M., Thiagarajan, G., Williams, A. D., Chellgren, B. W., Creamer, T. P., and Wetzel, R. (2006) J Mol Biol 355, 524-535
25. Stott, K., Blackburn, J. M., Butler, P. J., and Perutz, M. (1995) Proc Natl Acad Sci U S A 92, 6509-6513
26. Tanaka, M., Morishima, I., Akagi, T., Hashikawa, T., and Nukina, N. (2001) J Biol Chem 276, 45470-45475
27. Tanaka, M., Machida, Y., Nishikawa, Y., Akagi, T., Morishima, I., Hashikawa, T., Fujisawa, T., and Nukina, N. (2002) Biochemistry 41, 10277-10286
28. Bevivino, A. E., and Loll, P. J. (2001) Proc Natl Acad Sci U S A 98, 11955-11960
29. Chow, M. K., Paulson, H. L., and Bottomley, S. P. (2004) *J Mol Biol* **335**, 333-341
30. Chow, M. K., Ellisdon, A. M., Cabrita, L. D., and Bottomley, S. P. (2004) *J Biol Chem* **279**, 47643-47651
31. Kleywegt, G. J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K., and Jones, T. A. (1994) *Structure* **2**, 1241-1258
32. Gunasekaran, K., Hagler, A. T., and Gierasch, L. M. (2004) *Proteins* **54**, 179-194
33. Eyles, S. J., and Gierasch, L. M. (2000) *Journal of Molecular Biology* **301**, 737-747
34. Krishnan, V. V., Sukumar, M., Gierasch, L. M., and Cosman, M. (2000) *Biochemistry* **39**, 9119-9129
35. Clark, P. L., Weston, B. F., and Gierasch, L. M. (1998) *Fold Des* **3**, 401-412
36. Clark, P. L., Liu, Z. P., Rizo, J., and Gierasch, L. M. (1997) *Nat Struct Biol* **4**, 883-886
37. Thompson, J., Bratt, J., and Banaszak, L. (1995) *J Mol Biol* **252**, 433-446
38. Griffin, A., Adams, S. R., Jones, J., and Tsien, R. Y. (2000) *Methods Enzymol* **327**, 565-578
39. Ignatova, Z., and Gierasch, L. M. (2004) *Proc Natl Acad Sci U S A* **101**, 523-528
40. Muchowski, P. J., Schaffar, G., Sittler, A., Wanker, E. E., Hayer-Hartl, M. K., and Hartl, F. U. (2000) *Proc Natl Acad Sci U S A* **97**, 7841-7846
41. Ignatova, Z., and Gierasch, L. M. (2005) *Biochemistry* **44**, 7266-7274
42. Ferrone, F. (1999) *Methods Enzymol* **309**, 256-274
43. Thakur, A. K., and Wetzel, R. (2002) *Proc Natl Acad Sci U S A* **99**, 17014-17019
44. Oberg, K., Chrunyk, B. A., Wetzel, R., and Fink, A. L. (1994) *Biochemistry* **33**, 2628-2634
45. Harper, J. D., and Lansbury, P. T., Jr. (1997) *Annu Rev Biochem* **66**, 385-407
46. DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P., and Aronin, N. (1997) *Science* **277**, 1990-1993
47. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C. M. (1999) *Proc Natl Acad Sci U S A* **96**, 3590-3594
48. Andreu, J. M., and Timasheff, S. N. (1986) *Methods Enzymol* **130**, 47-59
49. Clark, P. L., Liu, Z. P., Zhang, J., and Gierasch, L. M. (1996) *Protein Sci* **5**, 1108-1117
50. Chen, Y. W., Stott, K., and Perutz, M. F. (1999) *Proc Natl Acad Sci U S A* **96**, 1257-1261
51. Masino, L., Kelly, G., Leonard, K., Trottier, Y., and Pastore, A. (2002) *FEBS Lett* **513**, 267-272
52. Clerico, E. M., Peisajovich, S. G., Ceolin, M., Ghiringhelli, P. D., and Ermacora, M. R. (2000) *Biochim Biophys Acta* **1476**, 203-218
53. Armen, R. S., Bernard, B. M., Day, R., Alonso, D. O., and Daggett, V. (2005) *Proc Natl Acad Sci U S A* **102**, 13433-13438
54. Khare, S. D., Ding, F., Gwanmesia, K. N., and Dokholyan, N. V. (2005) *PLoS Comput Biol* **1**, 230-235
55. Ross, C. A., Poirier, M. A., Wanker, E. E., and Amzel, M. (2003) *Proc Natl Acad Sci U S A* **100**, 1-3
56. Perutz, M. F., Pope, B. J., Owen, D., Wanker, E. E., and Scherzinger, E. (2002) *Proc Natl Acad Sci U S A* **99**, 5596-5600
57. Sharma, D., Shinchuk, L. M., Inouye, H., Wetzel, R., and Kirschner, D. A. (2005) *Proteins* **61**, 398-411
58. Wang, X., Vitalis, A., Wyczalkowski, M. A., and Pappu, R. V. (2005) *Proteins, in press* (published on line, DOI 1097-0134)
59. Radley, T. L., Markowska, A. I., Bettinger, B. T., Ha, J. H., and Loh, S. N. (2003) *J Mol Biol* **332**, 529-536
60. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) *Proc Natl Acad Sci U S A* **102**, 15400-15405
61. Zhang, X., Smith, D. L., Merin, A. B., Engemann, S., Russel, D. E., Roark, M., Washington, S. L., Maxwell, M. M., Marsh, J. L., Thompson, L. M., Wanker, E. E., Young, A. B., Housman, D. E., Bates, G. P., Sherman, M. Y., and Kazantsev, A. G. (2005) *Proc Natl Acad Sci U S A* **102**, 892-897
FOOTNOTES

* We thank Ulrich Hartl, Manajit Hayer-Hartl and Rick Morimoto for providing polyQ-containing plasmids, Nick Ranzette and Steven Sandler for the help with fluorescence microscopy, Bharati Vasudeva Rao for help with sample preparation for electron microscopy, and Ron Wetzel for his insightful comments on this work. We are grateful to Joanna Swain for her critical reading of the manuscript. This research was supported in part by NIH grant GM027616 to LMG. ZI holds a Heisenberg fellowship (IG 73/1-1).

‡ Present address:  Max-Planck-Institute of Biochemistry, Am Klopferspitz 18a, 82152 Martinsried by München, Germany.

Abbreviations used: CD, circular dichroism; CRABP I, cellular retinoic acid-binding protein I; α-CT; α-chymotrypsin; EDT, ethanedithiol; FlAsH, 4',5'-bis(1,3,2-dithioarsolan-2-yl)fluorescein; iLBP, intracellular lipid-binding protein; IPTG, isopropyl γ-D-thiogalactoside; HD, Huntington’s Disease; Htt, huntingtin protein; htt, gene encoding huntingtin; LB, Luria broth; RA, retinoic acid.

FIGURE LEGENDS

Fig. 1. Chimeras of huntingtin exon 1 and a model protein, tetra-Cys CRABP I, show length-dependent tendency to aggregate in a bacterial expression system. (A) Schematic diagram of the tetra-Cys CRABP I chimeras with huntingtin (Htt) exon 1. The tetra-cysteine motif introduced for FlAsH binding is shown in yellow, the polyQ tract of Htt exon 1 is depicted in black, and the flanking regions of the Htt exon 1 are shown in orange. The sequence of Htt exon 1 is:

MATLEKLMKAFESLKSF(Q)nPPPPPPPPPPQLPQQQPQPQPQPPPQPPPPPPPPPPGSAAVEELHP.

For the C-terminal chimeras n=20, 33, 40, 53, or 64, and for the N-terminal n=20 or 53. (B) Time course of in-cell expression of FlAsH-labeled tetra-Cys CRABP I constructs with C-terminal polyQ expansions followed at 530 nm. Cells were pre-loaded with FlAsH before IPTG induction of protein expression at time zero. Fluorescence of tetra-Cys CRABP I Htt20 is shown in red, of tetra-Cys CRABP I Htt33 in light blue, tetra-Cys CRABP I Htt40 in orange, tetra-Cys CRABP I Htt53 in blue and tetra-Cys CRABP I Htt64 in dark red; tetra-Cys CRABP I-expressing cells are shown in green for comparison. (C) Partitioning of protein between soluble (black bars), insoluble (light gray bars) and SDS-insoluble (dark gray bars) cellular fractions after 4 hours of expression of tetra-Cys CRABP Htt exon 1 chimeras. Note that tetra-Cys CRABP I and tetra-Cys Htt20 remain soluble through the time-course of expression.

Fig. 2. N- and C-terminal fusions of Htt exon 1 to tetra-Cys CRABP I show comparable polyQ length-dependent aggregation. (A) In-cell expression of FlAsH-labeled fusions following IPTG induction of protein expression at time zero. Fluorescence (at 530 nm; excitation 500 nm) of tetra-Cys CRABP I-expressing cells is shown in green, tetra-Cys CRABP I Htt20 in red, Htt20 tetra-Cys CRABP I in pink, tetra-Cys CRABP I Htt53 in blue and Htt53 tetra-Cys CRABP I in pink, tetra-Cys CRABP I Htt53 in light blue. Cell density (OD600) is shown in the open symbols in the same colors. (B) Partitioning of tetra-Cys CRABP Htt53 between the soluble (black bars), insoluble (light gray bars) and SDS-insoluble (dark gray bars) cellular fractions. (C) Fluorescence microscopy images of E. coli cells expressing tetra-Cys CRABP Htt53 showing the conversion of the E. coli cells into long filamentous arrays coincident with the appearance of SDS-insoluble aggregates. (D) Reduction of the expression of tetra-Cys CRABP I Htt53 by chloramphenicol treatment leads to the appearance of a lag phase (2.5 µg/ml, open blue circles; and 3.5 µg/ml, closed blue
triangles). For comparison, the time-courses of FlAsH fluorescence of tetra-Cys CRABP I Htt53 (blue closed circles) and tetra-Cys CRABP I Htt20 (red closed circles) are shown. (E) Fluorescent images of intact aggregates isolated from labeled cells (four hours post-induction) expressing tetra-Cys CRABP I Htt53. The scale bar is 1 μm. (F) Electron microscopic image of intact aggregates isolated from labeled cells (four hours post-induction) expressing tetra-Cys CRABP I Htt53. The scale bar is 200 nm.

**Fig. 3.** Analysis of the aggregation kinetics of tetra-Cys CRABP Htt53 in vitro at 37 °C. Aggregation kinetics were followed by FlAsH fluorescence of the labeled tetra-Cys CRABP I Htt53 (A), and by the decrease of the amount of the soluble monomer (see Materials and Methods) (B). The initial concentrations of labeled tetra-Cys CRABP I Htt53 in these experiments were 1 μM ( ), 2.5 μM ( ), 5 μM ( ), 10 μM ( ), and 20 μM ( ). Solid lines represent fits to sigmoidal curves. Tetra-Cys CRABP I Htt20 showed no aggregation under the same conditions (not shown). (C) Determination of the nucleus size based on the initial aggregation rates (panels A and B), following the approach of Ferrone (42). The open symbols and dashed lines correspond to data points from the aggregation kinetics followed by the FlAsH fluorescence, and the data points from the decrease of the monomer concentrations are presented as closed symbols fitted with a solid line. Note that linear fits yielded a slope of 2.8571 (dashed line) and 2.9017 (solid line) (equal to \( n^* + 2 \)), consistent with a monomeric nucleus. The regression coefficients were 0.983 and 0.989, respectively. (D) Tetra-Cys CRABP I Htt53 aggregation can be accelerated by addition of pre-formed seeds. Aggregation of 2.5 μM FlAsH-labeled tetra-Cys CRABP I Htt53 with and without pre-formed aggregates was monitored by FlAsH fluorescence. Fluorescence values in each experiment were normalized to exclude the contribution of the pre-labeled seed. Reactions contained no seed ( ), 1% ( ), 2.5% ( ) and 5% ( ) seed from the initial protein concentration in each experiment. (E) Determination of critical concentration. The values represent the concentration of the soluble monomer measured at the end points of aggregation (after 20 h) as a function of the initial tetra-Cys CRABP I Htt53 concentration. The critical concentration varies between 33 and 50 nM (0.7-1.1 ng/ml).

**Fig. 4.** Fusion of Htt exon 1 with an extended polyQ tract (53 residues) perturbs the structure of tetra-Cys CRABP I. Tryptophan fluorescence spectra of the Htt-chimeras with C-terminal (A) and N-terminal (B) exon 1. Note that Htt exon 1 lacks tryptophan residues. Retinoic acid is the natural ligand of CRABP I and binds to the central hydrophobic cavity, quenches the intrinsic Trp-fluorescence. Far-UV CD spectra of the Htt-fusions with C-terminal (C) and N-terminal (D) polyQ tracts. (E) α-Chymotrypsin susceptibility of Htt variants. The main proteolytic product with a molecular mass of 10 kDa originates from the tetra-Cys CRABP.

**Fig. 5.** Urea-induced unfolding of Htt chimeras. Tetra-Cys CRABP I is shown in green closed circles (fit shown as a solid green line), tetra-Cys CRABP I Htt20 in red closed circles (fit shown as a solid red line), Htt20 tetra-Cys CRABP in red open circles (fit shown as a dashed red line), tetra-Cys CRABP I Htt53 in blue closed circles (fit shown as a solid blue line), and Htt53 tetra-Cys CRABP in open blue circles (fit shown as a dashed blue line).
Table 1. Equilibrium stability of tetra-Cys CRABP I and fusions to Htt exon 1.

| Construct                | $\Delta G^\circ$ (kcal/mol) | $m_{eq}$ (kcal/mol M) | $C_m$ (M) |
|--------------------------|-------------------------------|-----------------------|-----------|
| tetra-Cys CRABP I        | 3.6±0.12                      | 2.16±0.10             | 1.73±0.07 |
| tetra-Cys CRABP Htt20 (C)| 3.73±0.13                     | 2.10±0.09             | 1.81±0.08 |
| tetra-Cys CRABP Htt53 (C)| 2.23±0.12*                   | 2.12±0.11             | 1.05±0.08 |
| tetra-Cys CRABP Htt20 (N)| 3.82±0.14                     | 2.11±0.11             | 1.83±0.07 |
| tetra-Cys CRABP Htt53 (N)| 2.34±0.12*                   | 2.09±0.10             | 1.12±0.11 |

* $\Delta G^\circ$ for these constructs was estimated by assuming reversibility of urea denaturation.
Figure 1

Ignatova and Giersch
Figure 2

Ignatova and Giersch
Figure 3

Ignatova and Gierasch
Figure 4

Ignatova and Gierasch
Figure 5

Ignatova and Gierasch
Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent \( \beta \) barrel protein
Zoya Ignatova and Lila M. Gierasch

J. Biol. Chem. published online March 8, 2006

Access the most updated version of this article at doi: 10.1074/jbc.M511523200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts