Biophysical Underpinnings of the Repeat Length Dependence of Polyglutamine Amyloid Formation*

Received for publication, January 23, 2014, and in revised form, February 19, 2014
Published, JBC Papers in Press, March 4, 2014, DOI 10.1074/jbc.C114.552943
Elizabeth Landrum and Ronald Wetzel
From the Department of Structural Biology and the Pittsburgh Institute for Neurodegenerative Diseases, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260

Background: The molecular basis for the repeat length dependence of disease risk in polyglutamine diseases is not known.

Results: We show here a nonlinear dependence of aggregation propensity versus polyglutamine repeat length that qualitatively mirrors how age of onset depends on repeat length.

Conclusion: This provides further support for a role for aggregation in these diseases.

Significance: This lends support to therapeutic strategies targeting aggregation.

There are now 10 expanded CAG repeat diseases in which both disease risk and age of onset are strongly dependent on the repeat length of the polyglutamine (polyQ) sequence in the disease protein. Large, polyQ-rich inclusions in patient brains and in cell and animal models are consistent with the involvement of polyQ aggregation in the disease mechanism. This possibility is reinforced by studies showing strong repeat length dependence to the aggregation process, qualitatively mirroring the repeat length dependence of disease risk. Our understanding of the underlying biophysical principles that mediate the repeat length dependence of aggregation, however, is far from complete. A previous study of simple polyQ peptides showed that $N^\alpha$, the size of the critical nucleus that controls onset of aggregation, decreases from unfavorable tetramer to favorable monomer over the range $Q_23$ to $Q_{26}$. These data, however, do not explain why, for all peptides exhibiting $N^\alpha \sim 1$, spontaneous aggregation rates continue to increase with increasing repeat length. Here we describe a novel kinetics analyses that maps out the nonlinear dependence with repeat length of a nucleation efficiency term that is likely related to aspects of nucleus structure. This trend accounts for why nucleus size increases to tetrameric at repeat lengths of $Q_{23}$ or below. Intriguingly, both aggregation and age of onset trend with repeat length in similar ways, exhibiting large changes per added Gln at low repeat lengths and small changes per added Gln at relatively long repeat lengths.

*This work was supported, in whole or in part, by National Institutes of Health Grants R01 AG019322 and R01 GM099718.

1 To whom correspondence should be addressed: Dept. of Structural Biology, University of Pittsburgh School of Medicine, Biomedical Sciences Tower 3, 3501 Fifth Ave., Pittsburgh, PA 15260, Tel.: 412-383-5271; Fax: 412-648-9008; E-mail: rnetzel@pitt.edu.

There are a large number of triplet repeat expansion diseases in abnormal human biology (1). Some pathogenic triplet expansions occur in noncoding DNA segments, whereas others occur in coding regions. One fascinating subset of the latter grouping are the 10 known expanded CAG repeat diseases that are associated with expansions of a polyglutamine (polyQ)2 sequence in a disease protein (2). These maladies, which include Huntington disease (3, 4), are progressive neurological disorders often presenting with movement or psychiatric symptoms (2). The boundaries between benign and pathogenic polyQ repeat lengths vary for different diseases from about 20 to about 45, with pathological thresholds in the range of 30–40 in most cases (2). In each disease, inheritance of a CAG repeat only slightly longer than the threshold confers a dramatic increase in disease risk. Above this characteristic threshold, age of onset decreases as repeat length increases according to a nonlinear trend featuring relatively large decreases in age of onset for repeat lengths just above the threshold, and much smaller decreases with each added CAG repeat at much longer repeat lengths (5, 6). Furthermore, age of onset and severity of symptoms actually decrease, in a mouse model of Huntington disease, at a very long polyQ repeat length (7). One possible explanation for these nonlinear trends in age of onset is an attenuation in the efficiency of formation of the toxic species at longer repeat lengths. Although the molecular mechanisms by which polyQ tract expansion triggers disease onset remain undefined (8–15), it seems reasonable to expect that, whatever the underlying molecular events that trigger the early cellular events in disease onset, these events should exhibit their own characteristic repeat length dependence.

One proposed family of molecular mechanisms is associated with the tendency of polyQ chains to aggregate into amyloid fibrils or other assemblies (16–18). Neuronal inclusions staining for polyQ have been identified in autopsy material in all 10 of the known expanded polyQ diseases, and polyQ aggregates are also a recurring feature of cell and animal models of polyQ diseases (2, 19, 20). Although evidence has been presented that the very large inclusions visible by light microscopy are more protective than toxic (21), it is very likely that expanded polyQ-producing cells harbor a variety of much smaller aggregates that are invisible to such techniques (22). For example, small oligomers and amyloid fibrils have been identified in studies of polyQ protein aggregation in vitro (23–26), oligomeric forms of unknown molecular structure can be detected in cells (27, 28) and in extracts of animal models (29), and isolated amyloid fibrils have been imaged in a cell culture model by super-resolution fluorescence microscopy (30). Importantly, polyQ aggregation both in vitro (31) and in cell (32) and animal (33) models.

2 The abbreviations used are: polyQ, polyglutamine; Aβ, amyloid-β.
exhibits a repeat length dependence that mirrors remarkably well the repeat length dependence of disease and age of onset.

Few studies have been reported examining the repeat length dependence of polyQ spontaneous amyloid formation and its underlying basis. One challenge in designing such studies is the fact that disease-associated polyQ sequences are always situated in the interior portions of the disease proteins (2), and, at least in some cases, the flanking sequences to the polyQ can have enormous impacts on their aggregation tendencies and mechanisms (25, 34). Nonetheless, the repeat length dependence of aggregation rate of polyQ sequences appears to be fundamental, manifesting not only in simple polyQ peptides with only a few solubilizing charged flanking residues (31) but also in the disease protein context (25). Given that polyQ is the common feature of all expanded polyQ disease proteins, it seems appropriate to first examine the structural basis for the repeat length effect in the isolated polyQ sequence.

Recently, we reported that one strong correlation with polyQ repeat length is the nucleus size, or critical nucleus, \(N^c\), for aggregation (35). We found that sequences containing Q23 or lower require a tetrameric nucleus (i.e. \(N^c = 4\)) for initiation of amyloid growth, whereas sequences of Q26 or above require formation of a monomeric nucleus (\(N^c = 1\)) that is presumed to be some specific, highly unfavorable folding state (35, 36). Because aggregation via a monomeric nucleus is much more efficient, sequences of Q26 or higher exhibit much faster spontaneous aggregation as compared with Q23 or shorter polyQ at similar concentrations (35). This dramatic change in nucleus size does not fully account for the repeat length dependence of aggregation, however. In particular, whereas all repeat lengths of Q26 or above initiate aggregation via a monomeric nucleus (35), previous studies suggest that aggregation rates continue to increase as repeat lengths increase above Q26 (31, 33, 37). Furthermore, the underlying explanation for why a change in nucleus size occurs precisely in the Q23–Q26 repeat length regime remains a complete mystery. Finally, it is worth considering that whereas repeat length clearly influences the kinetics of aggregation, there may also be a role for repeat length differences in aggregate stability in defining whether aggregates of particular polyQ repeat lengths can actually accumulate under in vivo conditions.

Here we conduct a quantitative analysis of the nucleation kinetics of a series of simple polyQ peptides with respect to polyQ repeat length, placing a special emphasis on a nucleation kinetics parameter not previously studied in a systematic way. We also investigate the dependence of fibril stability on repeat length. The results reveal clear patterns in both comparisons, suggesting that both may play a role in aggregation in vivo and hence perhaps in polyQ toxicity. The results also suggest a fundamental transition region for both properties in the Q23 to Q26 region, suggesting a common structural rationale.

**EXPERIMENTAL PROCEDURES**

All peptides were acquired crude from the Small Scale Synthesis facility at the Keck Biotechnology Resource Laboratory of Yale University. Peptides were purified and disaggregated, and reactions were prepared at 37 °C and monitored by HPLC as described (38, 39). Dissociation reactions to determine \(\Delta G_{\text{elong}}\) were performed by diluting end stage aggregation reactions with PBS and following dissociation upon incubation at 37 °C by HPLC (38). Nucleation kinetics analysis of the HPLC/sedimentation aggregation kinetics data was performed as described (38, 39). \(N^c\) values were calculated by subtracting 2 from the slope of the log-log plot of rate versus concentration (40). To obtain y-intercept values (log[0.5 k, \(n^2K_{\text{eq}}\)]) of the log-log plots for the data analysis described here, experimental data were extrapolated by applying the closest integer slope (i.e. 3 for \(N^c = 1\), etc.).

**RESULTS**

Based on their multimeric nature, aggregation rates typically exhibit a dependence on starting concentration, and details of this dependence can provide important information on the mechanisms of how reactions are initiated ("nucleation") (41) and how aggregates, once established, are propagated ("elongation") (42). Sigmoidal time courses are typically observed for nucleated growth polymerization reactions such as those undergone by simple polyQ peptides (41). A general mechanism for this type of process is shown in Fig. 1A. The slow onset of aggregation, often referred to as a lag time, reflects the operation of a high energy barrier to the organization of an aggregate nucleus (Fig. 1A, \(N^c\)). As is typical for aggregation dominated by primary nucleation (41), lag phases for early polyglutamine aggregation kinetics tend to exhibit rather gradual onsets of aggregation; this is in contrast to the abrupt onset of aggregation observed in cases of secondary nucleation (41, 43). Once initiated, aggregation reactions proceed until monomer is essentially depleted and a position of equilibrium (Fig. 1A, last line) is reached (44).

Incubation at 37 °C of freshly disaggregated (see "Experimental Procedures") K2Q42K2 in PBS leads to a time-dependent loss of monomer from solution that exhibits this kind of typical sigmoidal aggregation kinetics curve (Fig. 1B). Incubation at lower starting concentrations leads to similar sigmoidal time courses featuring longer lag times (Fig. 1B). All the peptides in this study exhibited analogous concentration-dependent, sigmoidal aggregation time courses. All of the aggregated products appear to be homogeneous, amyloid-like structures by electron microscopy (not shown).

Critical parameters of the nucleation mechanism can be extracted from the concentration-dependent kinetics data using the mathematical approach of Ferrone and co-workers (see "Experimental Procedures") (40, 41) based on a thermodynamic model of aggregation nucleation (Fig. 1A). The culmination of this analysis is a log-log plot of initial aggregation rate with respect to starting concentration. The slope of the log-log plot is related to the critical nucleus \(N^c\), which is the number of molecules required to form the nucleus. The y-intercept of the plot is related to the efficiency with which this critical nucleus promotes aggregation initiation, based on underlying parameters for the equilibrium constant for nucleus formation (\(K_{\text{eq}}\)) and the rate constants for elongation of the nucleus and growing fibril (\(k_{\text{fibr}}\)) (Fig. 1A).

The log-log plot for K2Q42K2 is shown in Fig. 1C. The parameters extracted for K2Q42K2, as well as for other simple polyQ peptides, are summarized in Table 1. The table shows that the
REPORT: Repeat Length Dependence of Polyglutamine Aggregation

FIGURE 1. Analysis of nucleated growth polymerization of K2QnK2 amyloid formation. A, nucleated growth polymerization mechanism (41) where M is the monomer ensemble, N* is the critical nucleus, F species are growing fibrils, $K_{Cr}$ is the equilibrium constant for nucleus formation, $k_{+}$ values are macroscopic elongation rate constants for the nucleus and nascent fibril, and $K_{r}$ is the association constant for fibril elongation. B, aggregation kinetics monitored by HPLC sedimentation assay at starting concentrations 131 μM (●), 97.4 μM (○), 58.1 μM (△), 31.3 μM (■), and 8.6 μM (■). C, log-log plot of initial aggregation rates versus initial concentrations. D, determination of $C_{r}$ value by monitoring dissociation of monomer from amyloid fibrils to an equilibrium concentration of monomer. E, $\Delta G_{\text{along}}$ values, derived from measured $C_{r}$ values, versus polyQ repeat length.

TABLE 1 Parameters for polyglutamine nucleation and elongation

All data are from this work, unless otherwise indicated.

| Repeat length | N* | $\log_{10}(0.5k_{+}^{-1}K_{r}^{-1})$ | $C_{r}$ |
|---------------|----|----------------------------------|--------|
| 18            | 3.7 | ND                     | 3.6 |
| 23            | 3.9 | ND                     | 3.0 |
| 25            | ND  | ND                     | 1.1  |
| 26            | 1.0 | -3.15                  | 1.0   |
| 26            | 0.9 | -3.33                  | ND    |
| 27            | 1.2 | -3.62                  | 0.8   |
| 27            | 0.9 | -3.08                  | ND    |
| 29            | 0.8 | -2.48                  | ND    |
| 30            | 0.7 | -2.80                  | ND    |
| 31            | 1.0 | -1.87                  | 0.8   |
| 34            | 1.0 | -1.68                  | ND    |
| 37            | 0.9 | -2.10                  | ND    |
| 42            | 0.5 | -0.92                  | 0.3   |
| 47            | 0.9 | -0.22                  | ND    |

N* values for aggregation of all peptides examined tend to be fractional values lying in the vicinity of 1, consistent with a monomeric nucleus. Similar values have been obtained previously for simple polyQ peptides with repeat lengths of Q26 or higher (35, 36, 40, 45–47). Although it is possible that genuine fractional values for N* might reflect subfibrils of the nucleation mechanism (48), it is important to note that the precision of the log-log plots, even when they include up to 10 data points, is moderate. Thus, although our typical experimental scatter-plots can reliably distinguish between N* = 1 and N* = 2, they cannot reliably distinguish between an N* of 1.0 and fractional values hovering around N* = 1 (35). Table 1 also shows the y-intercept values from the log-log plots. More positive values of the y-intercept are associated with greater efficiency of conversion of the critical nucleus into propagating fibrils. Thus, larger $K_{Cr}$ values favor the transient formation of nuclei in equilibrium with the monomer ensemble, and larger $k_{+}$ values favor more efficient conversion of these nuclei into growing fibrils that, once formed, are energetically unlikely to completely dissociate (Fig. 1A). Table 1 shows that for simple polyQ peptides with N* values in the range of 1, y-intercept values become more positive as repeat length increases.

Although the concentration of monomer when fibril formation reaches equilibrium (Fig. 1A, last line) is often only a very small percentage of the starting concentration, the final concentration is not zero and is actually highly significant. If the number of fibrils, and hence their molar concentration, is static as equilibrium is approached (Fig. 1A, last line), and if the free energy change for elongation is essentially independent of fibril length, then the concentration of monomers at equilibrium, $C_{r}$, is equivalent to the average fibril dissociation constant, $K_{d}$ (44). The inverse is therefore $K_{d}$ (Fig. 1A, last line), which has an associated free energy of elongation, $\Delta G_{\text{along}}$ (44). Thus, the lower the $C_{r}$ value, the farther the equilibrium position is to the right, and the more resistant the fibrils are to dissociation.

The derived $\Delta G_{\text{along}}$ can be remarkably robust, for example by giving information on the energetic costs of mutations for amyloid formation that are in very good agreement with the energetic costs of folding for the same mutations in a globular protein (49). The most robust way to determine the $C_{r}$ is by monitoring the dissociation of freshly made fibrils; a typical determination by this method is shown for K2Q42K2 in Fig. 1D. In contrast to amyloid fibrils of Aβ40, which dissociate to equilibrium within 1 day in PBS at 37 °C (44), K2Q42K2 and other polyQ fibrils take weeks to dissociate to equilibrium (Fig. 1D). It should be noted that although the best fit of the data in Fig. 1D is a sigmoidal curve, confidence is not high that this is a realistic representation of dissociation kinetics, which would require substantially more data points to determine. On the other hand, confidence in the plateau values (the $C_{r}$) of such plots is high because they generally match extremely well with
**DISCUSSION**

Our data (Table 1) show that there is a general improvement in nucleation efficiency (the $k_+^2 K_{\text{eq}}$ term) as well as amyloid fibril stability (the $C_r$ term) as polyQ repeat length increases. Beyond these general trends lie some subtleties that are worth exploring. For example, graphical representation of the $C_r$ data (Fig. 1E) shows a nonlinearity to the increase in fibril stability with increasing repeat length. Between Q18 and Q23, $\Delta G_{\text{elong}}$ changes by only 0.1 kJ/mol per residue, and between Q25 and Q27, the average change is about 0.17 kJ/mol per residue. However, between Q23 and Q25, $\Delta G_{\text{elong}}$ changes by about 1.3 kJ/mol per residue. This represents a dramatic increase in fibril stability over a very short polyQ repeat length range, suggesting an underlying structural cause. Interestingly, recent vibrational CD and hydrogen exchange studies also suggest dramatic changes in polyQ amyloid structure as polyQ repeat length increases over a similarly narrow range (51).

A qualitatively similar discontinuity was observed in the Q26 range for the trend of nucleation efficiency with repeat length, and we explored the basis of this nonlinearity in some detail. Fig. 2A shows a plot of $\log(0.5 k_+ K_{\text{eq}}^*)$, the $y$-intercept of the log-log plots, versus repeat length for peptides in the series K2QpK2 of various polyQ repeat length. Data shown are from this study (35) and (36) (●). Also shown are points representing the maximum possible values for the $y$ axis for K2QpK2 (●) and K2QpK2 (○) peptides aggregating via an $N^* = 1$ mechanism (see “Discussion” and panel B of this figure). The lines represent exponential fits to different sets of data points: dashed (●, ●); solid (●, ●). B, log-log plot estimating the maximal possible value for the $y$-intercept for a K2QpK2 peptide aggregating via an $N^* = 1$ mechanism. Concentration-dependent aggregation data (○) are from Ref. 35. The solid line is the best fit straight line to these data, giving a slope of 5.9 and an $N^*$ of 4.0, as described (35). The dashed line is a theoretical line of slope = 3.0 drawn through the lowest experimental data point and extrapolated to the $y$ axis. See “Discussion” for details. C, transformation of the panel A fit line showing how the $y$-intercept value changes with polyQ repeat length at different polyQ repeat lengths. D, experimental initial aggregation rates at different starting concentrations for the peptides K2QpGQpK2 (●) and K2QpGQpK2 (●), where $p = \ell$-Pro. The best fit straight lines give $N^*$ values of 0.81 and 0.98, respectively. Data were therefore fit to straight lines of slope 3.0, and the lines were extrapolated to the $y$ axis to give intercepts of 4.39 and 3.42, respectively. This gives a difference in $y$-intercepts of 0.98.

FIGURE 2. *Analysis of repeat length dependence of polyQ peptides aggregating by an $N^* = 1$ mechanism.* A, scatterplot of log(0.5 $k_+^2 K_{\text{eq}}^*$), the $y$-intercept of log-log plots, versus repeat length for peptides in the series K2QpK2 of various polyQ repeat length. Data shown are from this study (●) or from references (35) (●) and (36) (○). Also shown are points representing the maximum possible values for the $y$ axis for K2QpK2 (●) and K2QpK2 (○) peptides aggregating via an $N^* = 1$ mechanism (see “Discussion” and panel B of this figure). The lines represent exponential fits to different sets of data points: dashed (●, ●); solid (●, ●). B, log-log plot estimating the maximal possible value for the $y$-intercept for a K2QpK2 peptide aggregating via an $N^* = 1$ mechanism. Concentration-dependent aggregation data (○) are from Ref. 35. The solid line is the best fit straight line to these data, giving a slope of 5.9 and an $N^*$ of 4.0, as described (35). The dashed line is a theoretical line of slope = 3.0 drawn through the lowest experimental data point and extrapolated to the $y$ axis. See “Discussion” for details. C, transformation of the panel A fit line showing how the $y$-intercept value changes with polyQ repeat length at different polyQ repeat lengths. D, experimental initial aggregation rates at different starting concentrations for the peptides K2QpGQpK2 (●) and K2QpGQpK2 (●), where $p = \ell$-Pro. The best fit straight lines give $N^*$ values of 0.81 and 0.98, respectively. Data were therefore fit to straight lines of slope 3.0, and the lines were extrapolated to the $y$ axis to give intercepts of 4.39 and 3.42, respectively. This gives a difference in $y$-intercepts of 0.98.
REPORT: Repeat Length Dependence of Polyglutamine Aggregation

could be added to Fig. 2A. This is not practical, however, given the clear preference for the $N^* = 4$ mechanism with these short polyQ peptides. We therefore devised an alternative strategy in which we (a) assumed the hypothesis to be correct, (b) determined an approximate $y$-intercept for a $K_2Q_{23}K_2$ peptide based on this assumption, (c) included this estimated data point in a fit of the experimental $y$-intercept data, and (d) used independently obtained experimental data to test the accuracy of the resulting trend line in the low repeat length regime of the fit. [It is important to note that the $y$-intercept values for $K_2Q_{23}K_2$ aggregating by either an $N^* = 1$ or an $N^* = 4$ mechanism will be different because of the difference in the value of $K_{N^*}$ expected for the two mechanisms; thus, one cannot use the $y$-intercept for the $N^* = 4$ mechanism in the proposed analysis.]

Our approach is to inquire what the highest $y$-intercept value might be for an $N^* = 1$ mechanism for $K_2Q_{23}K_2$ such that it would still be cleanly outcompeted by the $N^* = 4$ mechanism and hence not observed experimentally. This was done using the graphical analysis shown in Fig. 2B. This figure shows the standard log-log plot of aggregation rates versus starting concentration with the data points (○) previously described for $K_2Q_{23}K_2$ (35). These data points fall in the experimentally accessible part of concentration space shown by the gray shading; aggregation rates for poorly aggregating $K_2Q_{23}K_2$ peptides at concentrations outside of this range are extremely difficult to measure accurately. The solid line shows the fit to the experimental data leading to the previous determination (35) that $N^* = 4$ under these conditions. The dashed line shows the hypothetical log-log fit, with a slope of 3 corresponding to $N^* = 4$ mechanism. The dashed line is drawn so that any data points on this line would represent slower aggregation rates than those experimentally observed (arrows). This indicates that an $N^* = 1$ mechanism for $K_2Q_{23}K_2$ might be observable were it not for the relative efficiency, in the experimental concentration range, of the $N^* = 4$ mechanism. The $y$-intercept of the dashed line, $-4.68$, is thus the highest possible $y$-intercept for an $N^* = 1$ mechanism for $K_2Q_{23}K_2$ such that it would still not be experimentally observed due to the competing $N^* = 4$ mechanism. We believe that the actual $y$-intercept value must be reasonably close to this $y$-intercept upper limit because $N^*$ is systematically increasing from 1 to 4 as repeat length shrinks from $Q_{26}$ to $Q_{23}$ (35).

The addition of this upper limit estimate for $K_2Q_{23}K_2$ (●) to Fig. 2A shows that it qualitatively continues the general trend of the experimental data (■▲◆). These data, plus the $K_2Q_{23}K_2$ estimate, yield a good fit to an exponential equation that emphasizes the nonlinear relationship in the data (Fig. 2A, solid line). Interestingly, when a graphical analysis (not shown) of previously obtained (35) $K_2Q_{14}K_2$ data is conducted in analogy to that shown in Fig. 2B, it yields an upper limit (○) $y$-intercept that falls well above the Fig. 2A solid trend line. This is consistent with the hypothesis that although the $N^* = 1$ and $N^* = 4$ mechanisms are of comparable efficiency in the $Q_{24}$–$Q_{25}$ repeat length range, as repeat lengths decrease further in this concentration range, the $N^* = 1$ mechanism becomes increasingly unfavorable and hence the $N^* = 4$ mechanism becomes increasingly favored.

The ever increasing challenge of kick-starting amyloid growth via a monomeric nucleation mechanism as polyQ repeat lengths become shorter is shown in Fig. 2C, which is a transform of the Fig. 2A solid line fit. It shows that the $y$-intercept value is predicted to change by only about 0.07 for each added or subtracted Gln residue in the $Q_{34}$ range, but is predicted to change by about 0.7, or 10 times as much, in the $Q_{16}$ range. Although the $\Delta y$/residue values predicted for polyQ in the $Q_{26}$–$Q_{47}$ range are based on experiment, the $\Delta y$/residue values in the $Q_{20}$ range cannot be experimentally tested directly for simple polyQ peptides in this concentration range because the $N^* = 1$ mechanism is not competitive with the $N^* = 4$ mechanism under such conditions. However, it is possible to devise an experimental test for the reasonableness of the Fig. 2A fit and the resulting Fig. 2C trend by constructing short polyQ analogs constrained, through introduction of $\beta$-hairpin-inducing mutations (52), to aggregate via an $N^* = 1$ mechanism. The log-log plots of two such peptides are shown in Fig. 2D. As expected, although both peptides aggregate via $N^*$ values in the range of 1, the longer peptide, $K_2Q_{10}pGQ_{10}K_2$ (a nominal $Q_{22}$ peptide counting the central D-Pro and Gly residues), aggregates significantly faster than the shorter peptide, $K_2Q_{23}pGQ_{23}K_2$ (a nominal $Q_{30}$ peptide). This is because $K_2Q_{10}pGQ_{10}K_2$ has a more favorable $y$-intercept, which was found to be 0.98 higher than the corresponding value for $K_2Q_{23}pGQ_{23}K_2$. This corresponds to a $\Delta y$/residue value of 0.49, which is in very good agreement with the $\Delta y$/residue value of 0.53 predicted for polyQ in the $Q_{23}$ range based on the curve in Fig. 2C. We believe that this agreement supports our underlying hypothesis and the methods used to test it.

We interpret the results shown in Fig. 2 as being consistent with the hypothesis that aggregation via a monomeric nucleus becomes increasingly less competitive as polyQ repeat length decreases until, in the $K_2Q_{23}K_2$ range, it becomes so unfavorable that it is out-competed by an alternative nucleation mechanism involving a tetrameric nucleus. This happens, despite the added constraint of diffusional entropy in assembling a tetrameric nucleus, because of the high inefficiency associated with productive nucleation via a monomeric nucleus. This high inefficiency at low polyQ repeat length might be due to a highly unfavorable $K_{N^*}$, highly unfavorable $k_+$, values, or both (Fig. 1A).

We think it is most likely that the solid trend line in Fig. 2A is primarily due to the impact of repeat length on the folding equilibrium constant for the monomeric nucleus, $K_{N^*}$. Steep dependence of the folding free energy on polyQ repeat length might be expected, for example, if the monomeric nucleus for polyQ aggregation is a highly unfavorable $\beta$-hairpin, as has been suggested (35, 45, 52). Although detailed repeat-length-dependent calculations are needed, our results are consistent with preliminary computational studies of polyQ folding and self-association (53, 54).

The graphical analysis shown in Fig. 2B also suggests that at starting concentrations lower than about 100 $\mu$M, the monomeric nucleation pathway will become more favorable than the tetrameric nucleation pathway for a $K_2Q_{23}K_2$ peptide. However, although this prediction may be quantitatively correct, the point is that aggregation by both mechanisms is so unfavorable...
for short polyQ sequences at such low concentrations as to be unobservable.

Together, our results provide new details for how polyQ repeat length influences the likelihood that amyloid aggregates will emerge and accumulate. Decreases in polyQ repeat length below Q26 affect the energetics of fibril formation especially dramatically. First, low polyQ repeat lengths destabilize the structures required for aggregation nucleation to such an extent that the monomeric nucleation mechanism that is operative at higher polyQ repeat lengths becomes poorly competitive. Second, lower repeat lengths lead to a dramatic reduction in fibril stability so that it becomes reasonable to consider that the poor thermodynamic stabilities of amyloid fibrils of short polyQs may well be responsible for the lack of amyloid accumulation by such sequences in cells and animals.

In contrast to the dramatic effects operating at the low end of the repeat length range examined here, we find very modest systematic increases in nucleation efficiency and fibril stability as repeat lengths increase above the Q26 range. Although we were not able to examine the very long repeat lengths found in some patients, the curvature in Fig. 2C suggests that the Δy/residue parameter will continue to get smaller and smaller as repeat lengths increase for polyQ sequences with higher repeats than those studied here. Such results could provide a mechanistic explanation for the significant curvature found in correlations of ages of onset versus CAG repeat length in several CAG repeat expansion diseases, such that in patients with very high repeat lengths the average decrease in age of onset for each added CAG is substantially smaller than the corresponding value for lower repeat lengths (5, 6).

Acknowledgment—We thank Karunakar Kar for assistance in acquiring some of the C values.

REFERENCES
1. van Eyk, C. L., and Richards, R. I. (2012) Dynamic mutations: where are they now? Adv. Exp. Med. Biol. 769, 55–77
2. Bates, G. P., and Benn, C. (2002) The polyglutamine diseases. in Huntington’s Disease (Bates, G. P., Harper, P. S., and Jones, L., eds), pp. 429–472, Oxford University Press, Oxford, UK
3. Bates, G., Harper, P. S., and Jones, L. (eds) (2002) Huntington’s Disease, Oxford University Press, Oxford, UK
4. Bates, G., Tabrizi, S. J., and Jones, L. (eds) (2014) Huntington’s Disease, Oxford University Press, Oxford, UK
5. Gusella, J. F., and MacDonald, M. E. (2000) Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. Nat. Rev. Neurosci. 1, 109–115
6. Andresen, J. M., Gayn, J., Djourssé, L., Roberts, S., Brocklebank, D., Chetty, S. S., US-Venezuela Collaborative Research Group, HD MAPS Collaborative Research Group, Cardon, L. R., Gusella, J. F., MacDonald, M. E., Myers, R. H., Houseman, D. E., and Wexler, N. S. (2007) The relationship between CAG repeat length and age of onset differs for Huntington’s disease patients with juvenile onset or adult onset. Ann. Hum. Genet. 71, 295–301
7. Dragatis, I., Goldowitz, D., Del Mar, N., Deng, Y. P., Meade, C. A., Liu, L., Sun, Z., Dietrich, P., Yue, J., and Reiner, A. (2009) CAG repeat lengths ≥335 attenuate the phenotype in the R6/2 Huntington’s disease transgenic mouse. Neurobiol. Dis. 33, 315–330
8. Zuccato, C., Valenza, M., and Cattaneo, E. (2010) Molecular mechanisms and potential therapeutic targets in Huntington’s disease. Physiol. Rev. 90, 905–981
9. Takahashi, T., Katada, S., and Onodera, O. (2010) Polyglutamine Diseases: Where does toxicity come from? What is toxicity? Where are we going? J. Mol. Cell Biol. 2, 180–191
10. Ross, C. A., and Tabrizi, S. J. (2011) Huntington’s disease: from molecular pathogenesis to clinical treatment. Lancet neurology 10, 83–98
11. Jimenez-Sanchez, M., Thomson, F., Zavodszky, E., and Rubinstein, D. C. (2012) Autophagy and polyglutamine diseases. Prog. Neurobiol. 97, 67–82
12. Orr, H. T. (2012) Polyglutamine neurodegeneration: expanded glutamines enhance native functions. Curr. Opin. Genet. Dev. 22, 251–255
13. Blum, E. S., Schwendeman, A. R., and Shaham, S. (2013) PolyQ disease: misfiring of a developmental cell death program? Trends Cell Biol. 23, 168–174
14. Labbadia, J., and Morimoto, R. I. (2013) Huntington’s disease: underlying molecular mechanisms and emerging concepts. Trends Biochem. Sci. 38, 378–385
15. Hughes, A., and Jones, L. (2014) Pathogenic mechanisms. in Huntington’s Disease (Bates, G., Tabrizi, S. J., and Jones, L. eds), 4th Ed., pp. 323–369, Oxford University Press, Oxford, UK
16. Bates, G. (2003) Huntington aggregation and toxicity in Huntington’s disease. Lancet 361, 1642–1644
17. Michałik, A., and Van Broeckhoven, C. (2003) Pathogenesis of polyglutamine disorders: aggregation revisited. Hum. Mol. Genet. 12, Suppl. 2, R173–R186
18. Davranche, A., Aviolat, H., Zeder-Lutz, G., Busso, D., Altschuh, D., Trottier, Y., and Klein, F. A. C. (2011) Huntington affinity for partners is not changed by polyglutamine length: aggregation itself triggers aberrant interactions. Hum. Mol. Genet. 20, 2795–2806
19. Rudnicki, D. D., Pletnikova, O., Vonsattel, J. P., Ross, C. A., and Margolis, R. L. (2008) A comparison of Huntington disease and Huntington disease-like 2 neuropathology. J. Neuropathol. Exp. Neurol. 67, 366–374
20. Wilburn, B., Rudnicki, D. D., Zhao, J., Weitz, T. M., Cheng, Y., Gu, X., Greiner, E., Park, C. S., Wang, N., Sopher, B. L., La Spada, A. R., Osmand, A., Margolis, R. L., Sun, Y. E., and Yang, X. W. (2011) An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington’s disease-like 2 mice. Neuron 70, 427–440
21. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R., and Finkbeiner, S. (2004) Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature 431, 805–810
22. Wetzel, R., and Mishra, R. (2014) Structural biology: Disorder, disorder, and conformational flux. in Huntington’s Disease (Bates, G., Tabrizi, S. J., and Jones, L. eds), pp. 274–322, Oxford University Press, Oxford, UK
23. Scherzinger, E., Lux, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrah, H., and Wanker, E. E. (1997) Huntington-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. Cell 90, 549–558
24. Poirier, M. A., Li, H., Macosko, J., Cai, S., Amzel, M., and Ross, C. A. (2002) Huntington spheroids and protofibrils as precursors in polyglutamine fibrilization. J. Biol. Chem. 277, 41032–41037
25. Thakur, A. K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V., M., Byeom, I. J., Anjum, D. H., Kodali, R., Creamer, T. P., Conway, J. F., Granenborn, A. M., and Wetzel, R. (2009) Polyglutamate disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. Nat. Struct. Mol. Biol. 16, 380–389
26. Legleiter, J., Mitchell, E., Lotz, G., G., Sapp, E., Ng, C., DiFiglia, M., Thompson, L. M., and Muchowski, P. J. (2010) Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner in vitro and in vivo. J. Biol. Chem. 285, 14777–14790
27. Takahashi, Y., Okamoto, Y., Popiel, H. A., Fujikake, N., Toda, T., Kinjo, M., and Nagai, Y. (2007) Detection of polyglutamine protein oligomers in cells by fluorescence correlation spectroscopy. J. Biol. Chem. 282, 24039–24048
28. Ottos, G., Digman, M. A., Aiken, C., Lukacsovich, T., Marsh, J. L., and Gratton, E. (2010) A two-step path to inclusion formation of huntingtin peptides revealed by number and brightness analysis. Biophys. J. 98, 3078–3085
29. Sathasivam, K., Lane, A., Legleiter, J., Warley, A., Woodman, B., Finkbeiner, S., Paganetti, P., Muchowski, P. J., Wilson, S., and Bates, G. P. (2010) Identical oligomeric and fibrillar structures captured from the
brains of R6/2 and knock-in mouse models of Huntington’s disease. *Hum. Mol. Genet.* 19, 65–78.
30. Sahl, S. J., Weiss, L. E., Duim, W. C., Frydman, J., and Moerner, W. E. (2012) Cellular inclusion bodies of mutant huntingtin exon 1 obscure small fibrillar aggregate species. *Sci. Rep.* 2, 895.
31. Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001) Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytoxicity. *J. Mol. Biol.* 311, 173–182.
32. Cooper, J. K., Schilling, G., Peters, M. F., Herring, W. J., Sharp, A. H., Kaminsky, Z., Masone, J., Khan, F. A., Delanoy, M., Borchelt, D. R., Dawson, V. L., Dawson, T. M., and Ross, C. A. (1998) Truncated N-terminal fragments of huntingtin with expanded glutamine repeats form nuclear and cytoplasmic aggregates in cell culture. *Hum. Mol. Genet.* 7, 783–790.
33. Morley, J. F., Brignull, H. R., Weyers, J. J., and Morimoto, R. I. (2002) The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10417–10422.
34. Ellisdon, A. M., Thomas, B., and Bottomley, S. P. (2006) The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J. Biol. Chem.* 281, 16888–16896.
35. Kar, K., Jayaraman, M., Sahoo, B., Kodali, R., and Wetzel, R. (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat. Struct. Mol. Biol.* 18, 328–336.
36. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) Polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15400–15405.
37. Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G. P., Lehraeh, H., and Wanker, E. (1999) Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington’s disease pathology. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4604–4609.
38. O’Nuallain, B., Thakur, A. K., Williams, A. D., Bhattacharrya, A. M., Chen, S., Thiagarajan, G., and Wetzel, R. (2006) Kinetics and thermodynamics of amyloid assembly using a high-performance liquid chromatography-based sedimentation assay. *Methods Enzymol.* 413, 34–74.
39. Jayaraman, M., Thakur, A. K., Kar, K., Kodali, R., and Wetzel, R. (2011) Assays for studying nucleated aggregation of polyglutamine proteins. *Methods* 53, 246–254.
40. Chen, S., Ferrone, F. A., and Wetzel, R. (2002) Huntington’s disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11884–11889.
41. Ferrone, F. (1999) Analysis of protein aggregation kinetics. *Methods Enzymol.* 309, 256–274.
42. Cannon, M. J., Williams, A. D., Wetzel, R., and Myszka, D. G. (2004) Kinetic analysis of β-amyloid fibril elongation. *Anal. Biochem.* 328, 67–75.
43. Knowles, T. P., Waudby, C. A., Devlin, G. L., Cohen, S. I., Aguzzi, A., Vendruscolo, M., Terentjev, E. M., Welland, M. E., and Dobson, C. M. (2009) An analytical solution to the kinetics of breakable filament assembly. *Science* 326, 1533–1537.
44. O’Nuallain, B., Shivaprasad, S., Kheterpal, I., and Wetzel, R. (2005) Thermodynamics of Aβ1–40 amyloid fibril elongation. *Biochemistry* 44, 12709–12718.
45. Thakur, A. K., and Wetzel, R. (2002) Mutational analysis of the structural organization of polyglutamine aggregates. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17014–17019.
46. Bhattacharyya, A., Thakur, A. K., Chellgren, V. M., Thiagarajan, G., Williams, A. D., Chellgren, B. W., Creamer, T. P., and Wetzel, R. (2006) Oligogoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* 355, 524–535.
47. Slepko, N., Bhattacharyya, A. M., Jackson, G. R., Steffan, J. S., Marsh, J. L., Thompson, L. M., and Wetzel, R. (2006) Normal-repeat-length polyglutamine peptides accelerate aggregation nucleation and cytotoxicity of expanded polyglutamine proteins. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14367–14372.
48. Vitalis, A., and Pappu, R. V. (2011) Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides. *Biophys. Chem.* 159, 14–23.
49. Williams, A. D., Shivaprasad, S., and Wetzel, R. (2006) Alanine scanning mutagenesis of Aβ(1–40) amyloid fibril stability. *J. Mol. Biol.* 357, 1283–1294.
50. Lee, J., Culyba, E. K., Powers, E. T., and Kelly, I. W. (2011) Amyloid-β forms fibrils by nucleated conformational conversion of oligomers. *Nat. Chem. Biol.* 7, 602–609.
51. Kurouski, D., Kar, K., Wetzel, R., Dukor, R. K., Lednev, I. K., and Nafie, L. A. (2013) Levels of supramolecular chirality of polyglutamine aggregates revealed by vibrational circular dichroism. *FEBS Lett.* 587, 1638–1643.
52. Kar, K., Hoop, C. L., Drombosky, K. W., Baker, M. A., Kodali, R., Ardunin, L., van der Wel, P. C., Horne, W. S., and Wetzel, R. (2013) β-Hairpin-mediated nucleation of polyglutamine amyloid formation. *J. Mol. Biol.* 425, 1183–1197.
53. Nakano, M., Ebina, K., and Tanaka, S. (2013) Study of the aggregation mechanism of polyglutamine peptides using replica exchange molecular dynamics simulations. *J. Mol. Model.* 19, 1627–1639.
54. Nakano, M., Watanabe, H., Rothstein, S. M., and Tanaka, S. (2010) Comparative characterization of short monomeric polyglutamine peptides by replica exchange molecular dynamics simulation. *J. Phys. Chem. B* 114, 7056–7061.