Published in final edited form as:

J Mol Biol. 2008 December 5; 384(1): 279–297. doi:10.1016/j.jmb.2008.09.026.

Atomistic Simulations of the Effects of Polyglutamine Chain Length and Solvent Quality on Conformational Equilibria and Spontaneous Homodimerization

Andreas Vitalis, Xiaoling Wang[†], and Rohit V. Pappu^{*}

Department of Biomedical Engineering and Center for Computational Biology, Washington University in St. Louis, One Brookings Drive, Campus Box 1097, St. Louis, MO 63130

Summary

Aggregation of expanded polyglutamine tracts is associated with nine different neurodegenerative diseases, including Huntington's disease. Experiments and computer simulations have demonstrated that monomeric forms of polyglutamine molecules sample heterogeneous sets of collapsed structures in water. The current work focuses on a mechanistic characterization of polyglutamine homodimerization as a function of chain length and temperature. These studies were carried out using molecular simulations based on a recently developed continuum solvation model that was designed for studying conformational and binding equilibria of intrinsically disordered molecules such as polyglutamine systems. The main results are as follows: Polyglutamine molecules form disordered, collapsed globules in aqueous solution. These molecules spontaneously associate at conditions approaching those of typical *in vitro* experiments for chains of length $N \geq 15$. The spontaneity of these homotypic associations increases with increasing chain length and decreases with increasing temperature. Similar and generic driving forces govern both collapse and spontaneous homodimerization of polyglutamine in aqueous milieu. Collapse and dimerization maximize self-interactions and reduce the interface between polyglutamine molecules and the surrounding solvent. Other than these generic considerations, there do not appear to be any specific structural requirements for either chain collapse or chain dimerization, *i.e.*, both collapse and dimerization are non-specific in that disordered globules form disordered dimers. In fact, it is shown that the driving force for intermolecular associations is governed by spontaneous conformational fluctuations within monomeric polyglutamine. These results suggest that polyglutamine aggregation is unlikely to follow a homogeneous nucleation mechanism with the monomer as the critical nucleus. Instead, the results support the formation of disordered, non beta-sheet-like soluble molten oligomers as early intermediates – a proposal that is congruent with recent experimental data.

Keywords

Aggregation; Coil-to-Globule Transitions; Homodimerization; Homogeneous nucleation; Polyglutamine

Introduction

Expanded polyglutamine tracts in unrelated proteins are responsible for nine different neurodegenerative diseases.¹ These include Huntington's disease (HD),² spinal and bulbar

^{*}Corresponding author Email address of corresponding author: pappu@wustl.edu.

[†]Current address: Pharmaceutical Sciences Research and Development, Global Biologics, Pfizer, Inc. BB581-D, 700 Chesterfield Pkwy West, Chesterfield, MO 63017

muscular atrophy, dentatorubral pallidoluysian atrophy,³ and the spinocerebellar ataxias 1,2,3,6,7,17.⁴ Each disease is associated with the deposition of neuronal intranuclear inclusions rich in aggregated polyglutamine.⁵ Ages of onset of these diseases are inversely correlated with the lengths of polyglutamine repeats.⁵ In HD, symptoms appear if the lengths of polyglutamine tracts exceed a threshold of 35–40 residues.^{6; 7; 8; 9}

Proteins with expanded polyglutamine repeats have increased susceptibility to proteolysis.^{10; 11; 12; 13; 14; 15; 16; 17} Products of proteolysis are rich in polyglutamine and these fragments can recruit glutamine-rich domains from other proteins and sequester them into growing aggregates.^{18; 19; 20; 21; 22; 23} Recruitment and sequestration can be toxic because it deprives cells of crucial proteins.²⁴ It can also be a source of chronic stress,²⁵ which in turn leads to an acceleration of the aging program by compromising protein homeostasis.^{26; 27; 28; 29; 30; 31; 32; 33} The length dependence of intrinsic rates and spontaneity of polyglutamine aggregation are of direct relevance to the recruitment and sequestration hypothesis for toxicity.^{34; 35; 36; 37}

In vitro biophysical studies have yielded important insights regarding conformational equilibria and aggregation of various polyglutamine constructs.^{13; 23; 38; 39; 40; 41; 42; 43; 44; 45} The main results are as follows: Monomeric polyglutamine-rich constructs are intrinsically disordered and this holds true irrespective of polyglutamine length.^{23; 46; 47; 48} The spontaneity and overall rate of aggregation increases with polyglutamine length.^{23; 41; 45} Synthetic peptides rich in polyglutamine form large aggregates with many morphological and dye-binding characteristics that mark these aggregates as being amyloid-like.^{13; 40; 49}

To obtain a definitive biophysical explanation for the observed length dependence of polyglutamine aggregation we need answers to two specific questions: First, why should polyglutamine molecules, which are polar, aggregate and/or be insoluble in aqueous milieus, *i.e.*, what are the driving forces for polyglutamine aggregation?^{50; 51; 52; 53; 54; 55} Second, what are the mechanisms by which polyglutamine molecules self-associate to form aggregates?⁴¹ Recent work from our lab has focused on answering the first question.^{52; 53; 54; 55} Here, we turn our attention to seek answers to the second question.

We proposed a connection between the phenomenology of polyglutamine aggregation and the well-established field of conformational and phase equilibria of synthetic homopolymers.^{53; 56} The relevant concepts have been reviewed recently⁵³ and originate in the seminal works of Flory,^{57; 58} Huggins,^{59; 60} and others.^{61; 62; 63; 64; 65; 66; 67; 68} For homopolymers there is a clear driving force for aggregation if polymers are in milieus that are so-called poor solvents. In such environments, polymers form homogeneously mixed solutions of isolated globules under dilute solution conditions. As concentration increases, we enter the two-phase regime where there is a clear driving force for phase separation/aggregation. In the single molecule limit, intra-chain interactions are preferred to chain-solvent interactions for chains in a poor solvent, and chain sizes measured using radii of gyration (R_g) or hydrodynamic radii (R_h) scale as $N^{1/3}$ with chain length.^{69; 70; 71} As a result, the conformational ensemble in dilute solutions comprises of compact, roughly spherical conformations, which minimize the interface between chain molecules and the surrounding solvent. In poor solvents, the stabilities of collapsed structures and the spontaneities of homotypic intermolecular associations will increase with chain length.^{53; 61; 62; 66; 67; 68; 72; 73; 74; 75}

Recent computational and fluorescence correlation spectroscopy studies helped establish that aqueous milieus at ca. 25°C are poor solvents for polyglutamine.^{52; 54; 55} Crick et al.⁵² quantified the scaling of R_h for monomeric polyglutamine in aqueous solutions and showed that chain size scales with chain length as $N^{1/3}$. Therefore, in aqueous milieus, individual polyglutamine molecules prefer collapsed structures that minimize interactions with the

surrounding solvent. The demonstration that water is a poor solvent for polyglutamine identifies a generic driving force for the aggregation of this system of molecules.⁶⁸

With knowledge of the driving force for aggregation, we turn to the question of whether polyglutamine aggregation follows homogeneous nucleation *i.e.*, is the formation of a specific, high energy, conformational species or a specific cluster of molecules a pre-requisite for aggregation?^{41; 76; 77; 78} The number of chain molecules within an aggregate can vary and the smallest aggregate is a dimer. If the mechanism of aggregation strictly follows homogeneous nucleation and dimerization is spontaneous, then the monomeric form of polyglutamine is the critical nucleus.⁴¹ In this scenario, the nucleus would be a specific, thermodynamically unfavorable, conformational species of the monomeric form. Alternatively, the nucleus size could be greater than or equal to two.⁷⁶ In this case dimerization along the productive aggregation pathway would not be spontaneous. Interpretations of experimental data lead to diametrically opposite views for the mechanism of polyglutamine aggregation.^{41; 42} Furthermore, theoretical work suggests that polymer aggregation in poor solvents does not follow homogeneous nucleation.⁶⁸ To adjudicate between conflicting suggestions, we need to interrogate intermolecular associations and conformational preferences realized at low concentrations and low copy numbers. This limit is inaccessible to most conventional assays for aggregation and hence we turn to computer simulations.

We present results from atomistic simulations of conformational equilibria and the monomer-dimer equilibrium for polyglutamine as a function of chain length and temperature. All of our simulations were carried out using the ABSINTH continuum solvation model that was developed with focus on simulating conformational equilibria and reversible associations of intrinsically disordered systems.⁷⁹ Details of the ABSINTH model and results from extensive validation of the Metropolis Monte Carlo simulation engine that uses the solvation model have been published recently.⁷⁹ Only the details relevant to the simulations of the polyglutamine system are presented in the methods section. Within the ABSINTH framework, conformational equilibria of various polypeptide systems change as a function of temperature, although the parameters of the model do not have explicit temperature dependencies. The ability to alter conformational equilibria using temperature as a control parameter, allows us to view temperature as a modulator of solvent quality.

The remainder of the text is organized as follows: We present results from quantitative studies of the length dependence of coil-to-globule transitions for monomeric polyglutamine molecules. This is followed by quantification of the length and temperature dependence of monomer-dimer equilibria. We then focus our analysis on the correlation between collapse and dimerization as well as the driving forces and conformational requirements for both processes. In the discussion section we summarize our results and place our findings in the context of the existing body of experimental data and proposals for mechanisms of polyglutamine aggregation.

Results

Length dependence of conformational equilibria for monomeric polyglutamine

We performed simulations for monomeric polyglutamine as a function of temperature and chain length. The peptide constructs in all of our simulations were of the form Ac-(Gln)_N-Nme. Here, Ac refers to the acetyl group and Nme denotes N-methylamide. For brevity, the sequences with *N* glutamine residues will hereafter be referred to as Q_N. Flexible polymers undergo coil-to-globule transitions that are akin to second order phase transitions and characterized by the existence of a “tri-critical” θ -point.^{69; 70; 71; 80; 81} At the θ -temperature ($T=T_\theta$), $\langle R_g \rangle$ is

proportional to $N^{0.5}$. Therefore, plots of $\xi = \frac{\langle R_g \rangle}{\sqrt{N}}$ as a function of temperature for different chain lengths should intersect at $T=T_\theta$. Theory also predicts that for $T>T_\theta$, the ratio ξ increases with increasing N , whereas for $T<T_\theta$ this ratio decreases as N increases. These predictions are consistent with the fact that for $T>T_\theta$, chain-solvent interactions are preferred in a so-called good solvent, the coil state is favored, and $\langle R_g \rangle$ scales as $N^{0.59}$. Conversely, for $T<T_\theta$, the chain collapses to minimize contacts with the poor solvent and $\langle R_g \rangle$ scales as $N^{0.33}$. Finally, as chain lengths increase, the sharpness of coil-to-globule transitions should increase, and the width of the transition region should decrease.

In Figure 1 we plot the variation of ξ as a function of simulation temperature for Q₅, Q₁₅, Q₃₀, and Q₄₅, respectively. We find that the coil-to-globule transition is ill-defined for the Q₅ peptide and this is consistent with the concept of “blobs”. Within a blob, the balance of chain-chain, chain-solvent, and relevant solvent-solvent interactions is smaller than $k_B T$. Here, T is temperature, and k_B is Boltzmann’s constant. If there are n residues in a blob, then the radius of gyration of the blob scales as $n^{1/2}$ and this scaling holds irrespective of solvent quality or temperature. From Figure 1 it is clear that Q₅ is essentially a blob-sized-peptide because ξ does not change significantly with temperature.

From Figure 1 we see that as chain lengths increase, the sharpness of coil-to-globule transitions increase and widths of transition regions decrease. The curves intersect at a common temperature of $T \approx 410\text{K}$. For temperatures that lie outside the transition region, $360\text{K} \leq T \leq 430\text{K}$, ξ decreases with increasing N in the globule limit ($T < 360\text{K}$) and it increases with increasing N in the coil limit ($T > 420\text{K}$). All of these observations are consistent with expectations listed above from the physics of generic, linear, flexible homopolymers. Such systems collapse in poor solvents in order to sequester themselves from unfavorable interactions with the surrounding milieu. In our calculations, $T < 360\text{K}$ corresponds to poor solvent conditions.

The results shown in Figure 1 suggest that $T \approx 410\text{K}$ is a reasonable estimate for T_θ . We test this proposal in Figure 2 where we plot the scaling of ensemble-averaged internal distances as a function of separation in linear sequence. At the theta point, ensemble averages of inter-residue distances $\langle R_{ij} \rangle$ scale as $|j-i|^{0.5}$ for $T=T_\theta$. Conversely, for $T < T_\theta$, especially if T is outside the transition region, $\langle R_{ij} \rangle$ for a range of sequence separations should plateau to a constant value proportional to the density of globules adopted in poor solvents.^{80; 81} Ensemble averaged internal distances $\langle R_{ij} \rangle$ were calculated as shown below:

$$\langle R_{ij} \rangle_T = \left\langle \frac{1}{n_{ij}} \sum_{k \in i} \sum_{l \in j} |\mathbf{r}_k^i - \mathbf{r}_l^j| \right\rangle_T \quad (1)$$

Here, $\mathbf{r}_k^i$ and $\mathbf{r}_l^j$ denote the position vectors of atoms k and l , which are part of residues i and j , respectively; n_{ij} denotes the number of unique pairwise distances between residues i and j and the angular brackets denote an average over all of our simulation data for the system in question at temperature T .

Figure 2 shows the variation of $\langle R_{ij} \rangle$ with sequence spacing $|j-i|$ for different chain lengths and temperatures. In panel (a) we see that $\langle R_{ij} \rangle$ increases systematically with sequence separation for Q₅ and this holds true irrespective of the simulation temperature. For longer chains and $T < 360\text{K}$, $\langle R_{ij} \rangle$ plateaus to fixed values for a range of sequence separations. This temperature regime mimics poor solvent conditions where collapsed states are preferred for monomeric polyglutamine. For $T > 360\text{K}$, the data in panels (b) – (d) show that $\langle R_{ij} \rangle$ increases

systematically with sequence separation and as T approaches 410K, $\langle R_{ij} \rangle$ scales as $|j-i|^{0.5}$ with $|j-i|$. This is demonstrated by favorable comparison of data at $T=410$ K for Q₁₅, Q₃₀, and Q₄₅ to dashed curves in panels (b)–(d) that plot $\langle R_{ij} \rangle$ as $R_0|j-i|^{0.5}$, where $R_0=5.7\text{\AA}$ is the value of $\langle R_{ij} \rangle$ for $|j-i| = 1$, for all chain lengths and temperatures.

Therefore, for the force field used in this work, $T=410$ K is a reasonable estimate for the θ -temperature for polyglutamine in aqueous solutions. At this temperature, polyglutamine molecules, specifically the longer chains ($N \geq 15$), are indifferent with regards to their preference for chain-chain versus chain-solvent interactions. The driving forces for intermolecular associations should be prominent below the θ -temperature. For $T > T_\theta$ or even for temperatures that are in the immediate vicinity of T_θ there is no *a priori* reason to expect favorable intermolecular associations because chain solvent interactions are favored to chain-chain interactions. We designate the θ -temperature as T_θ , the collapse temperature $T=360$ K as T_c , and the temperature $T=420$ K as T_s because swollen, random coil states are preferred for temperatures greater than T_s .

Length and temperature dependence of spontaneous homodimerization

We used replica exchange Metropolis Monte Carlo sampling to simulate homotypic associations of polyglutamine as a function of chain length and temperature. In addition to the monomer degrees of freedom, *i.e.*, backbone and sidechain torsion angles, rigid body degrees of freedom for each molecule were sampled. Details are presented in the methods section.

Figure 3 shows temperature dependent cumulative distribution functions $F(R)$ of intermolecular distances for pairs of Q₅, Q₁₅, Q₃₀, and Q₄₅ molecules. For a given pair of molecules, $F(R)$ is an estimate of the probability that the average intermolecular separation is less than or equal to R . For Q₅ the cumulative distribution functions are essentially independent of temperature. The likelihood of realizing a specific value of R increases with distance, suggesting that these molecules prefer to diffuse freely about each other. The conclusion is that both conformational equilibria and intermolecular associations for short glutamine-rich peptides are consistent with the behavior of short polar amides in water.

From the analysis of coil-to-globule transitions shown in Figure 1 we know that longer chains form more stable globules for $T < T_c=360$ K. This chain length dependent drive for intramolecular phase separation has consequences for the spontaneity of intermolecular associations as shown in three of the four panels of Figure 3. For temperatures in the range $T < 360$ K, the probability of spontaneous homodimerization increases with increasing chain length. For a given value of T , this is quantified in terms of higher probabilities associated with longer chains realizing close intermolecular separations. Conversely, for a given chain length, the probability of spontaneous dimerization decreases with increasing temperature. To quantify these observations, we computed excess interaction coefficients $B_{22}(T)$ using the cumulative distribution functions shown in Figure 3. These coefficients are defined as follows:

$$B_{22}(T) = \frac{\int_{R=0}^{R=R_{\text{droplet}}} [F_{T=T_\theta}(R) - F_T(R)] R^2 dR}{\int_{R=0}^{R=R_{\text{droplet}}} F_{T=T_\theta}(R) R^2 dR} \quad (2)$$

In equation (2), $F_T(R)$ is the cumulative distribution function at temperature T , $F_{T=T_\theta}(R)$ is the cumulative distribution function at T_θ , and $R_{\text{droplet}}=200\text{\AA}$ is the radius of the droplet used in the simulations (see methods section). The integrals were calculated using an extended trapezoidal rule. The excess interaction coefficients are in the spirit of normalized second virial coefficients that are routinely used in statistical thermodynamics to assess the magnitude of intermolecular associations in solutions of small molecules as well as flexible polymers.⁸²

⁸³ If $B_{22}(T)$ is less than zero, spontaneous homodimerization is thermodynamically favored vis-à-vis the θ -point and the degree of favorability is assessed by the magnitude of $B_{22}(T)$. If $B_{22}(T)$ is positive, then the chains avoid each other, more so than at the θ -point, indicating a clear preference for dissociated states. If the preference for associated and dissociated states is akin to that of an ideal chain, then $B_{22}(T)$ will be zero.

Figure 4 shows two sets of plots. Panel (a) plots the variation of $B_{22}(T)$ as a function of temperature for $T \leq 410\text{K}$. Separate curves are shown for each of Q_5 , Q_{15} , Q_{30} , and Q_{45} , respectively. For Q_5 , $B_{22}(T)$ is negligibly small across the entire temperature range; for the longer chains, $B_{22}(T)$ is negative over different temperature ranges and the magnitude of $B_{22}(T)$ decreases with increasing temperature. Specifically, $B_{22}(T)$ is negative in the temperature range $T \leq 315\text{K}$ for Q_{15} , and negative in the range $T \leq 360\text{K}$ for both Q_{30} and Q_{45} . Additionally, at temperatures where $B_{22}(T)$ is negative, its magnitude is greater for longer chains. This is summarized in panel (b) which plots the variation of $B_{22}(T)$ as a function of chain length and the different curves denote different temperatures.

The results from Figures 1, 3, and 4 may be summarized as follows: The sharpness of coil-to-globule transitions of monomeric polyglutamine increases with chain length. For $T \leq 360\text{K}$, Q_{30} and Q_{45} form stable globules, and in this temperature range these peptides form stable homodimers, whose stability decreases steadily with increasing temperature. For a given temperature in the range $T \leq 360\text{K}$, homodimers of Q_{45} are more stable than homodimers of Q_{30} . In contrast, homodimers of Q_{15} are only marginally stable and are accessible over a narrower temperature range $T \leq 315\text{K}$ and this weak dimerization is consistent with the shallow coil-to-globule transition observed for this system. The observation of length-dependent dimerization shows that the driving force for polyglutamine aggregation increases with chain length. We now analyze the physical basis for the length and temperature dependence of spontaneous homodimerization in polyglutamine.

Correlation between properties of monomeric polyglutamine and $B_{22}(T)$

Panels (a) and (b) of Figure 5 show correlations between the temperature dependencies of specific conformational characteristics of monomeric polyglutamine chains and the temperature dependence of $B_{22}(T)$ in the range $T \leq 360\text{K}$. Panel (a) shows a correlation between the density of monomeric globular polyglutamine and $B_{22}(T)$. The magnitude of the latter decreases as density decreases. Similarly, panel (b) shows that the magnitude of $B_{22}(T)$ decreases as the ensemble averaged value of R_g increases for monomeric globular polyglutamine. In both cases, the magnitudes of Pearson correlation coefficients that assess the degree of linear correlation are ca. 0.95.

For a given temperature, the driving forces for chain collapse may be decomposed into two components.⁸⁴ Specifically, the mean-field internal energy per residue may be written as:

$$\frac{\langle U \rangle}{N} = C_1(T) + C_2(T)N^{-1/3} \quad (3)$$

Here, N denotes chain length, $\langle U \rangle$ is the average potential energy at temperature T , $C_1(T)$ measures the bulk energy density, and $C_2(T)$ measures the surface energy density. $C_1(T)$ provides an estimate of the effective strength of self-interactions and $C_2(T)$ estimates the energy associated with making interfaces between polyglutamine and the surrounding solvent. Self-interactions are favorable if $C_1(T)$ is negative and the strengths of favorable interactions are measured by the magnitude of $C_1(T)$. If $C_2(T)$ is positive, its magnitude measures the energy penalty associated with increasing the size of the unfavorable chain-solvent interface. Conversely, if $C_2(T)$ is negative, then mixing of the chain and solvent is preferred.

Panels (a) and (b) in Figure S1 of the supplementary material plot the temperature dependence of $C_1(T)$ and $C_2(T)$ in the range $T \leq 360\text{K}$. The stabilities of collapsed states increase as $C_1(T)$ becomes more negative and $C_2(T)$ becomes more positive and this happens as temperature decreases in the range $T \leq 360\text{K}$. Comparisons of data in Figures 4 and 5 to those shown in Figure S1 indicate that $B_{22}(T)$ becomes increasingly more negative as the driving force (magnitudes of $C_1(T)$ and $C_2(T)$, respectively) for forming compact (small ζ) dense globules (large ρ) increases. Upon collapse, self-interactions are maximized. Dimerization leads to a diminution of the unfavorable solute-solvent interface and increased self-interactions through intermolecular association.

Since collapse and dimerization result from the combined drive to minimize unfavorable solute-solvent interfaces and maximize self-interactions, the surface-to-volume ratio (R_{SV}) of a single chain in a poor solvent provides a generic measure of the relevant driving forces.⁸⁵ For globules, R_{SV} decreases with increasing chain length because it scales as $N^{-1/3}$. For small N , such as Q_{15} the contribution from the surface term to R_{SV} is significant, which means that unfavorable surface energies are not readily offset by favorable self-interactions. As a result, Q_{15} shows weak tendencies toward collapse and stable dimerization. R_{SV} decreases rapidly for larger N . For Q_{30} and increasingly so for Q_{45} , R_{SV} is small enough to allow for the unfavorable surface energies to be offset by favorable self-interactions as evidenced by the increased preference for collapsed and dimerized states in the regime $T < 360\text{K}$. Q_{30} and Q_{45} encompass the threshold length range for polyglutamine disease phenotypes. The preceding discussion identifies R_{SV} as a rather simple measure for the observed phenotype and is anchored in generic concepts drawn from polymer physics. However, if polymers were rigid spheres or colloidal particles that interacted primarily through surface contacts,⁸⁶ then the value of R_{SV} for a single chain would be less meaningful because $C_1(T)$ would be negligible. In this case, the relevant surface-to-volume ratio would be that of clusters of molecules and not that of a single molecule. The description of polyglutamine aggregation would then follow classical models for homogeneous nucleation, where aggregation is favorable only if the cluster size is greater than some critical number.

Conformational specificity in collapse and intermolecular associations

The foregoing analysis focused on generic polymer physics parameters and the correlations between these quantities and the spontaneity of intermolecular associations. We also assessed the presence of specific, ensemble-averaged conformational propensities that can be implicated in promoting both collapse and intermolecular associations. Specifically, we asked if there is a discernible increase in β -sheet propensity associated with collapse, spontaneous associations, or both? To answer this question, we computed the fractional α -helical and β -sheet contents using our simulation data.

There are several ways to assess secondary structure content in proteins and polypeptides. We have developed a strategy that is based on analysis of distributions of backbone ϕ and ψ angles. The resultant measure, shown below, provides a reasonable estimate of secondary structure content as compared to popular measures such as DSSP⁸⁷ and STRIDE⁸⁸ (data not shown). The fractional α and β contents f_α and f_β , respectively are defined as:

$$f_x = \frac{1}{N} \sum_{i=1}^N f_x^{(i)}, \quad X \equiv \alpha \text{ or } \beta$$

$$f_x^{(i)} = \begin{cases} 1.0 & \text{if } (\varphi_i, \psi_i) \in X \\ \exp(-\tau_x d_{x(i)}^2) & \text{otherwise} \end{cases}$$

$$d_{x(i)}^2 = \left\{ \left(\sqrt{[(\varphi_i - \varphi_x) \bmod 2\pi]^2 + [(\psi_i - \psi_x) \bmod 2\pi]^2} - r_x \right) \bmod 2\pi \right\}^2 \quad (4)$$

In equation (4), f_X denotes the fractional content of secondary structure type X, where X is either α or β . The $\text{mod}2\pi$ terms corrects for periodicity effects when calculating distances in angular space and N is the number of residues in the sequence, excluding capping groups. The coordinates (ϕ_X, ψ_X) define the reference ϕ, ψ values to be adopted by an individual residue for the secondary structure motif of type X. If a residue i adopts ϕ and ψ angles that lie within a circle of radius r , then the parameter $f_X^{(i)}$ is set to unity; otherwise $f_X^{(i)}$ assumes a value between 0 and 1, and the precise value is determined by two parameters *viz.*, the distance $d_{X(i)}$ and τ_X . The latter is the width of the Gaussian function used to determine the value to be assigned for $f_X^{(i)}$. For $X \equiv \alpha$, $(\phi_\alpha, \psi_\alpha) = (-60^\circ, -50^\circ)$, $r_\alpha = 30^\circ$, and $\tau_\alpha = 0.002 \text{deg}^{-2}$. Conversely, for $X \equiv \beta$, $(\phi_\beta, \psi_\beta) = (-125^\circ, 125^\circ)$, $r_\beta = 40^\circ$, and $\tau_\beta = 0.002 \text{deg}^{-2}$.

Figure 6 shows four panels that summarize the temperature dependencies of fractional α -helical (f_α) and β -sheet (f_β) contents in Q₅, Q₁₅, Q₃₀, and Q₄₅. Data are shown from simulations of monomeric polyglutamine and those with two chains (“dimer”). From the data shown in Figure 6 we conclude the following: 1) There is clear, statistically significant diminution in f_α with increasing chain length. Inasmuch as intermolecular associations become favorable with increasing chain length, the decreased α -helical propensities with increasing N suggest a weak correlation between decreased helical propensity and chain associations. 2) While collapse and intermolecular associations show clear temperature dependencies, conformational propensities measured in terms of f_α and f_β show very weak temperature dependencies. Therefore, local conformational propensities appear to be only weakly coupled from the driving forces for collapse and intermolecular associations. 3) For a given solvent quality (defined by the value of the simulation temperature T), the fractional β content is greater than or equal to the fractional α content, and this is true irrespective of chain length. 4) Of the three chains, Q₁₅, Q₃₀, and Q₄₅, the two longer chains show a diminution in the α -helical content and an enhancement in β content by a weak increase in f_β with increasing temperature.

Local secondary structure content changes weakly as a function of temperature and chain length. Despite this, the driving forces for collapse and the spontaneities of homodimerization show clear temperature and length dependencies. Therefore, we conclude that disordered globules associate to form disordered dimers. This observation is congruent with the findings of Krishnan and Lindquist⁸⁹ for the aggregation of the NM regions of the yeast prion protein Sup35. They found that “molten oligomers”, which form as precursors to NM fiber formation, are dominated by contacts between the globular forms of the glutamine and asparagine rich N-domain,⁹⁰ which also forms collapsed structures in its monomeric form.⁹¹

Evidence for intrinsic disorder in polyglutamine

Experimental data and computational studies have documented the lack of conformational specificity in monomeric polyglutamine. In our simulation data, this preference for intrinsic disorder prevails despite the preference for collapsed states for temperatures in the range $T \leq 360\text{K}$. In previous work, we proposed that intrinsic disorder is a direct consequence of the homopolymeric nature of polyglutamine. The lack of sequence specificity implies that a variety of compact species, irrespective of chain conformation, have equivalent stabilities, and that the conformational ensemble therefore is a heterogeneous collection of compact conformations. While this type of disorder is distinct from the disorder associated with denatured proteins, collapse does not imply folding.⁹² Our analysis of local conformational propensities makes this point. Additionally, we can analyze the variations in contact patterns between individual members of the conformational ensemble to assess the degree of disorder as a function of temperature. To accomplish this, we quantify disorder by computing a single figure of merit namely, the normalized variance in the number of intramolecular contacts

(σ_N^2) as a function of temperature. This quantity, computed for monomeric polyglutamine, is defined as follows:

$$\sigma_N^2(T) = \frac{1}{NT} \sum_{k=1}^{n_{\max}(N)} \left(k - \langle n_c \rangle_T^{(N)} \right)^2 p_k^{(N)}(T)$$

$$\langle n_c \rangle_T^{(N)} = \sum_{k=1}^{n_{\max}(N)} k p_k^{(N)}(T)$$
(5)

Here, N denotes the chain length; T is the simulation temperature; $p_k^{(N)}(T)$ is the probability of realizing k intramolecular contacts in a chain of length N at temperature T ; a contact is defined by any two non-bonded atoms from residues i and j having a distance $\leq 3\text{\AA}$; $\langle n_c \rangle_T^{(N)}$ is the average number of intramolecular contacts in a chain of length N at temperature T ; and $n_{\max}(N)$ is the maximal number of realizable intramolecular contacts in a chain of length N .

Results for the variation of σ_N^2 as a function of temperature for Q₁₅, Q₃₀, and Q₄₅ are shown in panel (a) of Figure 7. In the high temperature limit ($T > 450\text{K}$), past the theta point, chains sample canonical denatured state ensembles where the dominant contacts are local and the likelihood of realizing distal contacts is very small; this is true for a majority of conformations in the ensemble. Consequently, the average number of contacts is small and so is the variance. Just below the theta point, the chains are in the transition region and sample conformations from two distinct ensembles *viz.*, the coil and globule states. In this regime, conformational fluctuations are large, and values for σ_N^2 are high because vastly different types of conformations are sampled. In complete congruence with the analysis shown in Figure 1, the sharpness of the coil-to-globule transition increases with N . This is manifest by the fact that as N increases, the width of the transition region decreases and the peak height increases in panel (a) of Figure 7. In the collapse regime, σ_N^2 decreases with decreasing temperature and does not plateau to a well-defined value. This systematic decrease of σ_N^2 with decreasing temperature is a characteristic signature of dynamical disorder and is consistent with glassy behavior quantified in previous studies of monomeric polyglutamine.⁵⁴ The assignment of dynamical disorder to collapsed polyglutamine is made clear by comparing the variance profiles shown in panel (a) of Figure 7 to the variance profiles obtained from simulation data for thermal unfolding of two well-folded proteins (shown in panel (b) of Figure 7) namely, the B1 domain of protein G (GB1) and the engrailed homeodomain (ENH). There are well-defined baselines in the values of σ_N^2 on either side of the transition region. Additionally, the unfolded baseline (high temperature) is higher in value than the folded baseline (low temperature), which is consistent with the adoption of a roughly rigid structure with small-scale fluctuations at low temperature and a heterogeneous ensemble characterized by dominant local contacts at high temperature. In contrast, for polyglutamine, there are temperatures well into the collapsed regime ($T < 360\text{K}$) for which σ_N^2 is actually higher than the asymptotic value achieved in the high temperature regime ($T > 450\text{K}$). These data are consistent with the proposal that monomeric polyglutamine fluctuates between disparate collections of conformations of roughly equivalent compactness. Intrinsic disorder results because collapse is only weakly coupled from folding in these simple systems that lack the requisite sequence specificity to prefer a specific compact conformation.

Importance of spontaneous fluctuations for promoting intermolecular associations

We have established that monomeric polyglutamine, which is intrinsically disordered, associates to form disordered homodimers. The latter point is underscored in the analysis where

we showed that local conformational propensities are essentially unchanged between the isolated monomer and associated dimer, Figure 6. Using a simple approach, we interrogated the role of intrinsic disorder (spontaneous fluctuations) of polyglutamine in promoting intermolecular associations. This was done in a series of simulations where we quantified the likelihood of realizing spontaneous associations of rigid globules. These simulations were carried out as follows: Random globular conformations were chosen from the conformational ensemble of monomeric Q₃₀ at $T=298\text{K}$. The internal coordinates were then frozen and only rigid body Monte Carlo moves were allowed for subsequent sampling. Statistics were recorded to construct the requisite histograms for intermolecular separations sampled in simulations with rigid globules. The process was repeated approximately thousand times, and the resultant, average cumulative distribution $F(R)$ was compared to that obtained for the association of “fully flexible” chains. These comparisons are shown in Figure 8. The dashed curve, which corresponds to the cumulative distribution function for rigid globules, reveals the importance of conformational fluctuations in promoting intermolecular associations.

The suppression of conformational fluctuations at $T=298\text{K}$ leads to a diminution of intermolecular associativity for Q₃₀. The lack of rigid structural preferences or a stable fold upon collapse is clearly responsible for promoting intermolecular associations between disordered globules. This result is consistent with the observation that many aggregation sequences are also intrinsically disordered. However, some caution is required in interpreting the results of Figure 8. For instance, Figure 7 shows that the degree of disorder measured by σ_N^2 increases with temperature for $T < 360\text{K}$ and yet $B_{22}(T)$ decreases with increasing temperature. The physical basis for the latter observation comes from the analysis in Figures 5 and S1, which demonstrates that the poorness of solvent decreases with increasing temperature. Therefore, we conclude that both poorness of solvent and spontaneous conformational fluctuations, work together to promote spontaneous homodimerization. The preference for homodimerization requires an appropriate combination of the poorness of solvent (see Figures 1–5) and magnitude of fluctuations. This point is reinforced by the following observation: In the temperature regime $T_c \leq T \leq T_\theta$ ($360\text{K} < T < 410\text{K}$) the magnitudes of conformational fluctuations go through a maximum for all chain lengths. In this regime, the surface energy penalty, measured by $C_2(T)$, is still positive and approaches zero only as $T \rightarrow T_\theta$ (data not shown). This implies homodimerization is thermodynamically unfavorable despite the fact the solute-solvent interface is also unfavorable. Under these conditions, homodimerization might require the formation of an appropriate conformational nucleus, to which only appropriate conformations would be able to dock and minimize the unfavorable interface with the surrounding solvent. Alternatively, some other, higher order, oligomeric species might be the thermodynamically favored entity because such a species might minimize the unfavorable solute-solvent interface more efficiently than dimers in the regime $T_c \leq T \leq T_\theta$. A detailed investigation of the precise correlation between poorness of solvent and the magnitude of conformational fluctuations merits further scrutiny and is reserved for a separate study.

Contacts that promote collapse and dimerization

Polyglutamine molecules are polyamides built by a repetition of backbone secondary amides and sidechain primary amides. To analyze the types of inter-atomic contacts that lead to collapse and dimerization, we computed site-site pair correlation functions. Results are shown in Figures S2 and S3 of the supplementary material. The former shows the intramolecular site-site correlations for Q₄₅ whereas the latter shows the intermolecular correlation functions. Both sets of correlation functions were constructed using simulation data with two Q₄₅ molecules. The relevant pair correlation functions are calculated with respect to T_θ , as described below.

The site-site correlation functions of interest to us are between backbone donors (N) and backbone acceptors (O), sidechain donors (N) and sidechain acceptors (O), backbone donors (N) and sidechain acceptors (O), and sidechain donors (N) and backbone acceptors (O). If we denote donor atoms as D and acceptor atoms as A, then the relevant donor-acceptor site-site correlation function $g_{DA}(r)$ at temperature T is computed as:

$$g_{DA}(r) = \frac{h_{DA}^{(T)}(r)}{h_{DA}^{(\theta)}(r)} \quad (6)$$

Here, $h_{DA}^{(T)}(r)$ is the histogram of relevant donor-acceptor distances at temperature T and $h_{DA}^{(\theta)}(r)$ is the corresponding histogram of distances at the theta point. If $g_{DA}(r) > 1$, then there is an enhancement of the relevant donor-acceptor contacts in the ensemble at temperature T vis-à-vis the theta point; if $g_{DA}(r) = 1$, then the distribution of donor-acceptor contacts at separation r is equivalent to that of the theta point; finally, if $g_{DA}(r) < 1$, then there is a depletion of donor-acceptor contacts at separation r vis-à-vis the theta point.

Figure S2 (supplementary material) shows a panel of intramolecular donor-acceptor site-site correlation functions for monomeric Q₄₅. As T approaches the theta point, all pair correlation functions converge upon values of unity for all distances. For lower temperatures, specifically $T \leq 360\text{K}$, there is significant enhancement vis-à-vis the theta point of short-range ($3\text{\AA} \leq r < 5\text{\AA}$) sidechain donor – sidechain acceptor and sidechain donor – backbone acceptor contacts. All four sets of site-site correlation functions in Figure S2 show systematic enhancements of medium-range contacts ($5\text{\AA} \leq r \leq 20\text{\AA}$) and depletion of distal contacts. This feature is consistent with the preference for collapsed states at lower temperatures. The pair correlation functions shown in Figure S2 suggest that the collapsed states are characterized by prominent sidechain-backbone interactions, again with respect to the theta point, indicating that the sidechain amides solvate the backbone, thereby minimizing the interface between backbone secondary amides and the aqueous milieu. The approach used here to calculate pair correlations differs from the one used in previous work.⁵⁴ Here we used the theta point as our reference state whereas in previous work we used an ideal chain model as the reference and therefore the two sets of correlation functions for collapsed, monomeric polyglutamine are different.

Figure S3 (supplementary material) shows intermolecular donor-acceptor site-site correlation functions calculated using simulation data for a pair of Q₄₅ molecules in the simulation volume. These pair correlation functions are shown on a log scale to facilitate the visualization of all the data. For temperatures that are in the collapse regime, there is significant enhancement of all flavors of donor-acceptor contacts. This is because of significant intermolecular donor-acceptor contacts that are absent at the theta point. These observations suggest that spontaneous dimerization of polyglutamine is the result of the drive to minimize the interface with the surrounding aqueous environment, a poor solvent for polyglutamine, and to replace this interface with favorable intra- and intermolecular contacts between all combinations of backbone donors, backbone acceptors, sidechain donors, and sidechain acceptors.

Discussion

Molecular simulations have played an important role in generating insights and testable hypotheses for various self-assembly phenomena involving folded proteins and intrinsically disordered proteins.^{93; 94; 95; 96; 97; 98; 99; 100; 101} The latter are challenging systems for simulation and experiment alike because their free energy landscapes are both degenerate and rugged.⁵⁴ Energy functions for simulations have multiple flavors. These range from structure-based energy functions to the molecular mechanics paradigm. In the former, the focus is on

the role of geometry in guiding folding and assembly. The latter rests on the premise of transferability whereby energy functions are designed by transferring parameters that accurately describe the conformational and phase equilibria of small model compounds.

The sequence simplicity of polyglutamine systems is attractive for testing predictions from a large body of theoretical work on synthetic homopolymers. Accordingly, several groups have carried out simulations of varying complexity to test specific ideas and hypotheses regarding polyglutamine conformational equilibria, changes associated with increasing chain length, and assembly mechanisms.^{102; 103; 104; 105; 106; 107} Structure-based Go models and other heuristic coarse graining approaches have been used to construct phase diagrams for polyglutamine as a function of chain length and numbers of molecules. The works of the Hall^{102; 103; 105} and Dokholyan^{108; 109} groups are particularly noteworthy. In their work, the chain length dependent likelihood of β -sheet formation is an important consideration in developing their energy functions and a mechanistic explanation for polyglutamine aggregation. Our ABSINTH continuum solvation model rests on the molecular mechanics paradigm and in this work we have presented results from atomistic simulations on the length and temperature dependence of conformational equilibria and spontaneous dimerization of polyglutamine molecules. Our main findings are as follows:

1. Profiles for coil-to-globule transitions (Figure 1) show a striking length dependence that is in good agreement with expectations from polymer physics theories. Specifically, the sharpness of this transition increases with N and this implies that the stability of collapsed states in aqueous milieus increases with chain length. Curves that plot the variation of normalized chain size with temperature coincide at the theta point, $T_\theta \approx 410\text{K}$. We find three distinct temperature regimes namely, $T < T_c$, where globules are favored, $T_c \leq T_\theta \leq T_s$ the transition region, and $T > T_s$, where swollen random coils are favored.
2. In the collapse regime, $T < T_c$, homodimerization is spontaneous. For a given temperature in the collapse regime, the spontaneity of homodimerization increases with chain length and conversely, for a given chain length, the spontaneity of homodimerization decreases with increasing temperature.
3. There is a tight correlation between the driving forces for collapse of monomeric polyglutamine and the spontaneity of homodimerization (Figures 5 and S1).
4. Driving forces for collapse of monomeric polyglutamine have two generic components. These are: (i) The drive to maximize self-interactions; and (ii) The drive to minimize the unfavorable solute-solvent interface with the surrounding aqueous milieu. The favorable, length and temperature dependent homodimerization we observe for polyglutamine may also be explained in terms of similar, generic driving forces. Dimerization should lead the formation of additional self-interactions (Figure S3) and the number of favorable intra- and intermolecular self-interactions should increase with chain length similar to the “linear lattice” effect proposed by Bennett et al.¹¹⁰ Similarly, dimerization leads to further minimization of the unfavorable solute-solvent interface.
5. Evidence that the length dependence of spontaneous homodimerization is non-specific and originates in generic considerations for polymers in poor solvents comes from Figure 6. Here, we showed that there are no substantial local conformational changes, such as β -sheet formation or conversion from α -helical forms that can be implicated in promoting homodimerization.
6. We show, in complete agreement with experimental data,²³ that polyglutamine, molecules irrespective of chain length, are intrinsically disordered even under conditions where collapsed states are thermodynamically favored. This intrinsic

disorder namely, the inability to adopt a stable fold can be implicated in the spontaneity of homodimerization. Suppression of disorder by quenching conformational fluctuations leads to a significant diminution in the preference for associated states.

7. Generic attributes of polymers in poor solvents (see 4 above) and spontaneous conformational fluctuations together give rise to a clear drive for non-specific homodimerization. Both requirements work in concert to promote dimerization. The precise interplay between solvent quality and conformational fluctuations merits further scrutiny.
8. We have shown that homodimerization is spontaneous and that the spontaneity increases with chain length for a prescribed poorness of the solvent. However, dimerization does not require the obligate formation of a specific, thermodynamically unfavorable, conformational species of the monomeric form. Our results also suggest that higher order oligomers can form readily, although this needs to be studied carefully. Hence, homogeneous nucleation is unlikely to be the correct mechanistic explanation for polyglutamine aggregation.

Connection to interpretations of experimental results

Chen et al.⁴¹ described the formation of large ordered polyglutamine aggregates as a nucleation-dependent reaction. They used the thermodynamic nucleus model of Ferrone, a variant of homogeneous nucleation, to analyze kinetic data for the formation of large aggregates. In the schematic that emerged, monomeric polyglutamine (irrespective of chain length) is in rapid pre-equilibrium with an ordered nucleus (presumably an ordered β -sheet conformation). The pre-equilibrium constant increases with increasing chain length, but remains immeasurably small for all chain lengths. According to this model, the lag phase in polyglutamine aggregation arises because β -sheet formation is thermodynamically unfavorable. However, this unfavorable folding is a conformational pre-requisite for the formation of aggregates of all sizes, including dimers.

Chen et al.⁴¹ used four probes to monitor the kinetics of aggregation. These were circular dichroism (CD), dynamic light scattering, Thioflavin T (ThT) binding, and reverse phase HPLC. If β -sheet contents do not vary with oligomerization, then CD signals would not change as oligomers formed. Similarly, dynamic light scattering cannot resolve the presence of smaller species; ThT binding most likely reports only on the formation of large ordered aggregates; and if the oligomers are part of the soluble species, then reverse phase HPLC would not detect these oligomers. Hence, close scrutiny of the data presented by Chen et al. suggests that one cannot rule out the presence of soluble oligomers in the reacting mixture.

Interestingly, other lines of experimental evidence support the presence of oligomers as identifiable intermediates.^{27; 111; 112; 113} Lee et al.⁴² recently measured the aggregation kinetics of Q₂₃ using peptide constructs that were similar to those used by Chen et al.⁴⁰ Using both static and dynamic light scattering, Lee et al. found evidence for the formation of soluble, linear aggregates during the lag phase. They also found the early aggregates to be lacking in regular secondary structure. Inasmuch as we can connect dimer formation with formation of larger aggregates, we propose that our results, which show a lack of local conformational specificity in chain collapse and intermolecular interactions, are consistent with the observations of Lee et al.

In comparison to other computational results, the studies by Marchut and Hall^{102; 103; 105} are the most relevant. They employed a conceptually different, structurally guided, coarse grain model to describe peptide and solvent. Despite this difference from our approach, a brief comparison of the results is in order. In their most recent study the concentration used is 2.5

mM for a system comprised of 24 molecules with chain lengths ranging from 16–48 residues at various reduced temperatures. Marchut and Hall find that at temperatures close to the effective T_0 of their model, large-scale aggregates with relatively large fractions of β -sheet hydrogen bonds and with distinctive ring-like topologies are populated. The authors note that experimental evidence for these structural motifs is lacking. At lower reduced temperatures, however, they describe “amorphous aggregates”, *i.e.*, aggregates lacking in structure. Our results are congruent with this unstructured “phase”. In their parlance, disordered globules with unstructured interfaces are termed amorphous aggregates. As for the observed ordered phase (“sheets”), we argue that the concentration regime in their work as well as the employed model predispose the results toward this order. This is suggested by the fact that β -rich structures appear even at the monomer level, which is incongruent with existing experimental data for monomeric polyglutamine. It is therefore likely that the appearance of an ordered phase for small oligomers is an artifact of the way their models were built. These differences notwithstanding, the results of both studies appear to share some overlap in predicting amorphous aggregates under certain conditions. Whether more of the results will be reconciled if the simulation conditions between the two studies are fully matched, remains to be seen, and is a topic for future investigation. Recently, Cecchini et al.¹¹⁴ used a surface area based implicit solvent model and replica exchange molecular dynamics to study the aggregation propensities of four short sequences including a 7-residue polyglutamine peptide. They found that the polyglutamine peptide and a peptide excised from the sequence of Sup35 showed clear aggregation propensities as discerned by analysis of the temperature dependence of a nematic order parameter. In future work we intend to reanalyze the current data using order parameters developed by the Caflisch group since these seem to provide an indicator of sequence regions that will have greater propensity for aggregation.

The presence of disordered low and high molecular weight aggregates implies that simple homogeneous nucleation models cannot describe polyglutamine aggregation. Questions persist regarding the degree of complexity needed in mechanistic models for polyglutamine aggregation. One possibility is that the formation of disordered linear or spherical aggregates, unlike the formation of ordered aggregates, occurs off pathway and does not follow the tenets of homogeneous nucleation theory. Alternatively, as suggested by Lee et al.⁴² and others^{89; 113; 115; 116}, disordered aggregates that are sufficiently large might convert to ordered forms. The different scenarios are summarized in Figure 9. The schematic in Figure 9 is congruent with the tenets of the generalized Lumry-Eyring model put forward recently by Andrews and Roberts.¹¹⁷ At this juncture we cannot adjudicate between the different possibilities shown in Figure 9. This will require simulations that explicitly probe the extent of coupling between specific conformational restraints and intermolecular associations. It will also require advances in computational methods that allow us to study conformational equilibria and aggregation of multiple molecules in atomic detail. From an experimental standpoint, we will need to probe the distribution of oligomers formed at low concentrations and low copy numbers because these will provide a direct way to compare results from simulations to experimental data. These investigations will require novel experiments or the adaptation of experiments used on other aggregation-prone systems to study polyglutamine aggregation.

Materials and Methods

System setup

Capped polypeptides composed exclusively of glutamine residues (Acetyl-(Gln) $_N$ -N-Methylamide) were built with fixed bond lengths and angles according to the Engh-Huber high-resolution, crystallographic geometries. Peptides with $N = 5, 15, 30$, and 45 were simulated in the $nNVT$ ensemble in a spherical droplet of radius $R_{\text{droplet}} = 200 \text{ \AA}$, where n denotes the number of individual molecules, N the number of glutamine residues, V the volume of the

simulation droplet, and T the simulation temperature. In all cases, polar interactions were truncated at 14Å, and short-range interactions were truncated at 10Å. This is justified by the fact that the simulation system does not include any naked charges and the charge set used, which is based on the OPLS-AA/L parameters, models primary (sidechain) and secondary (backbone) amides as dipoles, albeit with the partial charge model.

The simulation system consisted of either a single chain ($n = 1$) for the monomer simulations or of two chains ($n = 2$) for the dimer case. The chain molecules were confined to the simulation volume by applying a stiff harmonic boundary potential restraining the molecules to the simulation droplet. It is important to note that for a given simulation, fluctuations in n are quenched in the nVT ensemble. Therefore, the monomer simulations mimic the infinite dilution limit despite the fact that the simulation volume is finite. For the dimer case, the effective concentration was ca. 100 μM , which is in the concentration range of most *in vitro* experiments. Given our simulation setup, there is no possibility of studying the formation of oligomers larger than a dimer. This does not mean that we predict the absence of such larger species in the concentration range of 100 μM . Rather, our simulations focus on quantifying the spontaneities of chain-length and temperature dependent intermolecular associations at low copy numbers and high effective concentrations. This scenario might be reminiscent of *in vivo* settings, although this claim is purely speculative in nature. Extensions to study the formation of larger oligomers will require improvements in sampling methodologies and these are currently being pursued.

Conformational Sampling

All the simulations presented in this work were performed using Metropolis Monte Carlo (MMC) sampling of the relevant degrees of freedom, which for polyglutamine are the ϕ , ψ , and ω angles of the polypeptide backbone as well as the three sidechain dihedral angles (χ_1 , χ_2 , χ_3) of the glutamine residue. Additionally, for $n=2$, we include the sampling of rigid body degrees of freedom namely, translations of centers-of-mass and rotational reorientations of molecules. Details of the move sets employed are summarized in Table S1 of the supplementary material and all of our choices are explained in detail in the caption that accompanies this table. It is important to remind the reader that appropriate design of MMC move sets allow us to simultaneously probe multiple, disparate length scales rather efficiently, taking advantage of the low overall density. This situation is unlike molecular dynamics sampling, which is quite inefficient for sampling large-scale conformational changes as well as intermolecular associations/dissociations. The latter is hindered by slow diffusion and will require adaptive approaches and indeed MMC sampling may be viewed as a variant of such an adaptive approach. All simulations were carried out using our ABSINTH software package, which is being readied for public release (<http://lima.wustl.edu/absinth.php>), and are summarized in Table 1.

As was shown in previous work, intrinsically disordered polyglutamine systems pose a serious challenge for conformational sampling. To improve the quality of our simulation data, we used thermal replica exchange^{118; 119; 120}, which includes an additional Markov chain to the MMC sampling. Details of all the parameters for replica exchange are summarized in Table 1. While there is no formal requirement to restrict swaps between adjacent replicas, this restriction is implemented in practice because acceptance of proposed swaps between nonadjacent replicas is rather small. This restriction increases the overall efficiency of replica exchange given the finite number of swaps that are feasible during a simulation.^{118; 119; 120} In our simulations, we allowed swaps between all unique pairs of replicas¹²¹ in the range $298\text{K} \leq T \leq 410\text{K}$ because the acceptance of proposed swaps between nonadjacent replicas remained finite and we found that this improved the overall quality of sampling, especially for the lower temperatures.

Molecular mechanics force field

All of the data presented here were generated using the ABSINTH continuum solvation model, the details of which were published recently. In this model, the effects of solvent water are approximated by two contributions: 1) A direct mean-field interaction; and 2) A dielectric screening effect. The ABSINTH model⁷⁹, which is based on the EEF1 paradigm of Lazaridis and Karplus¹²², has been shown to reproduce the polymeric behavior of polyglutamine when compared to both simulations in explicit solvent as well as to experimental data. The results presented in the current work were obtained using two minor modifications to the previously published force field. First, for reasons of computational efficiency, partial charges on net-neutral methyl and methylene groups were omitted. All other partial charges are identical to those reported previously and are based on the OPLS-AA/L force field. Second, we employed slightly modified Lennard-Jones (LJ) parameters compared to the parameters published recently. The LJ parameters used in this work are shown in Table S2 (supplementary material). These modifications were not necessary; instead they are chronological artifacts because the simulations presented here were carried out using a previous, unpublished version of the ABSINTH force field and reflect the fact that the simulations presented here were completed prior to the fine-tuning of the published ABSINTH model. In light of the computational expense involved in carrying out the current simulations, we decided against repeating the calculations with the refined parameters. Brief tests (data not shown) indicate that all of the conclusions regarding the length and temperature dependencies of polyglutamine conformational equilibria and the spontaneities of homodimerization remain qualitatively robust. However, T_θ shifts to a value that is lower than 410K when we use the parameters that were published recently.

Data analysis

Most analysis quantities were computed once every 10^3 to 10^4 steps depending on the total extent of the simulation (see Table 1). For each monomer/dimer simulation, we carried out multiple, independent simulations with replica exchange. Therefore, we used a modified block averaging technique to estimate error bars. In this approach, for each temperature point, the data obtained from a single replica exchange simulation run is treated as a single block. With this approach, we are not confronted with the problem of having to chop our simulation data into *ad hoc* blocks. Independence of blocks for averaging is guaranteed because the starting conformations for all simulations are completely randomized. However, we do not have the resources to carry out hundreds of independent replica exchange runs. Instead, we typically have data from four independent replica exchange runs. Hence, the error bars that result from “block averaging”, which in reality is averaging over completely independent trajectories, are not rigorous estimates of statistical and bias errors in sampling; rather, they act as qualitative indicators of the reproducibility of our results between independent runs with randomized starting conformations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grant 5R01NS056114-01 from the National Institutes of Health. We thank Scott Crick, Carl Frieden, Nicholas Lyle, and Timothy Williamson for helpful discussions and suggestions. RVP is grateful to Ron Wetzel for insights and guidance during the initial stages of this work.

References

1. Zoghbi HY, Orr HT. Glutamine repeats and neurodegeneration. *Annu Rev Neurosci* 2000;23:217–247. [PubMed: 10845064]

2. MacDonald ME, Gusella JF. Huntington's disease: Translating a CAG repeat into a pathogenic mechanism. *Curr Opin Neurobiol* 1996;6:638–643. [PubMed: 8937828]
3. Becher MW, Rubinsztein DC, Leggo J, Wagster MV, Stine OC, Ranen NG, Franz ML, Abbott MH, Sherr M, MacMillan JC, Barron L, Porteous M, Harper PS, Ross CA. Dentatorubral and pallidoluysian atrophy (DRPLA) -Clinical and neuropathological findings in genetically confirmed North American and European pedigrees. *Mov Disord* 1997;12:519–530. [PubMed: 9251070]
4. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu Rev Neurosci* 2007;30:575–621. [PubMed: 17417937]
5. Morley JF, Brignull HR, Weyers JJ, Morimoto RI. 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* 2002;99:10417–10422. [PubMed: 12122205]
6. Macmillan JC, Snell RG, Tyler A, Houlihan GD, Fenton I, Cheadle JP, Lazarou LP, Shaw DJ, Harper PS. Molecular analysis and clinical correlations of the Huntington's-disease mutation. *Lancet* 1993;342:954–958. [PubMed: 8105214]
7. Verhoef S, Harper PS. Infantile Huntingtons Disease with an unusually large CAG expansion. *J Med Genet* 1997;34:553–553. [PubMed: 9222962]
8. MacDonald ME, Gines S, Gusella JF, Wheeler VC. Huntington's disease. *NeuroMol Med* 2003;4:7–20.
9. Penney JB, Vonsattel JP, MacDonald ME, Gusella JF, Myers RH. CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann Neurol* 1997;41:689–692. [PubMed: 9153534]
10. Waelter S, Boeddrich A, Lurz R, Scherzinger E, Lueder G, Lehrach H, Wanker EE. Accumulation of mutant huntingtin fragments in aggresome-like inclusion bodies as a result of insufficient protein degradation. *Mol Biol Cell* 2001;12:1393–1407. [PubMed: 11359930]
11. Thomas P, Wilkinson F, Man NT, Harper PS, Neal JW, Morris GE, Jones AL. Full length huntingtin is not detected in intranuclear inclusions in Huntington's disease brain. *Biochem Soc Trans* 1998;26:S243–S243. [PubMed: 9765962]
12. Krobitsch S, Lindquist S. Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. *Proc Natl Acad Sci U S A* 2000;97:1589–1594. [PubMed: 10677504]
13. Bader R, Seeliger MA, Kelly SE, Ilag LL, Meersman F, Limones A, Luisi BF, Dobson CM, Itzhaki LS. Folding and fibril formation of the cell cycle protein CksI. *J Biol Chem* 2006;281:18816–18824. [PubMed: 16675442]
14. Holmberg CI, Staniszewski KE, Mensah KN, Matouschek A, Morimoto RI. Inefficient degradation of truncated polyglutamine proteins by the proteasome. *EMBO J* 2004;23:4307–4318. [PubMed: 15470501]
15. Venkatraman P, Wetzel R, Tanaka M, Nukina N, Goldberg AL. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol Cell* 2004;14:95–104. [PubMed: 15068806]
16. Sawa A, Nagata E, Sutcliffe S, Dulloor P, Cascio MB, Ozeki Y, Roy S, Ross CA, Snyder SH. Huntingtin is cleaved by caspases in the cytoplasm and translocated to the nucleus via perinuclear sites in Huntington's disease patient lymphoblasts. *Neurobiol Dis* 2005;20:267–274. [PubMed: 15890517]
17. Sun B, Fan W, Balciunas A, Cooper JK, Bitan G, Steavenson S, Denis PE, Young Y, Adler B, Daugherty L, Manoukian R, Elliott G, Shen WY, Talvenheimo J, Teplow DB, Haniu M, Haldankar R, Wypych J, Ross CA, Citron M, Richards WG. Polyglutamine repeat length-dependent proteolysis of huntingtin. *Neurobiol Dis* 2002;11:111–122. [PubMed: 12460551]
18. Perez MK, Paulson HL, Pendse SJ, Saionz SJ, Bonini NM, Pittman RN. Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J Cell Biol* 1998;143:1457–1470. [PubMed: 9852144]
19. McCampbell A, Taylor JP, Taye AA, Robitschek J, Li M, Walcott J, Merry D, Chai YH, Paulson H, Sobue G, Fischbeck KH. CREB-binding protein sequestration by expanded polyglutamine. *Hum Mol Genet* 2000;9:2197–2202. [PubMed: 10958659]

20. Slepko N, Bhattacharyya AM, Jackson GR, Steffan JS, Marsh JL, Thompson LM, Wetzel R. Normal-repeat-length polyglutamine peptides accelerate aggregation nucleation and cytotoxicity of expanded polyglutamine proteins. *Proc Natl Acad Sci U S A* 2006;103:14367–14372. [PubMed: 16980414]
21. Duennwald ML, Jagadish S, Giorgini F, Muchowski PJ, Lindquist S. A network of protein interactions determines polyglutamine toxicity. *Proc Natl Acad Sci U S A* 2006;103:11051–11056. [PubMed: 16832049]
22. Matsumoto G, Kim S, Morimoto RI. Huntingtin and mutant SOD1 form aggregate structures with distinct molecular properties in human cells. *J Biol Chem* 2006;281:4477–4485. [PubMed: 16371362]
23. Chen S, Berthelie V, Yang W, Wetzel R. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J Mol Biol* 2001;311:173–182. [PubMed: 11469866]
24. Wetzel R. Nucleation of huntingtin aggregation in cells. *Nat Chem Biol* 2006;2:297–298. [PubMed: 16710335]
25. Brignull HR, Morley JF, Morimoto RI. The stress of misfolded proteins: C-elegans models for neurodegenerative disease and aging. *Molecular Aspects of the Stress Response: Chaperones, Membranes and Networks* 2007;594:167–189.
26. Satyal SH, Schmidt E, Kitagawa K, Sondheimer N, Lindquist S, Kramer JM, Morimoto RI. Polyglutamine aggregates alter protein folding homeostasis in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 2000;97:5750–5755. [PubMed: 10811890]
27. Wacker JL, Zareie MH, Fong H, Sarikaya M, Muchowski PJ. Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer. *Nat Struct Mol Biol* 2004;11:1215–1222. [PubMed: 15543156]
28. Miller VM, Nelson RF, Gouvion CM, Williams A, Rodriguez-Lebron E, Harper SQ, Davidson BL, Rebagliati MR, Paulson HL. CHIP suppresses polyglutamine aggregation and toxicity in vitro and in vivo. *J Neurosci* 2005;25:9152–9161. [PubMed: 16207874]
29. Cashikar AG, Duennwald M, Lindquist SL. A chaperone pathway in protein disaggregation - Hsp26 alters the nature of protein aggregates to facilitate reactivation by Hsp104. *J Biol Chem* 2005;280:23869–23875. [PubMed: 15845535]
30. Boeddrich A, Gaumer B, Haacke A, Tzvetkov N, Albrecht M, Evert BO, Muller EC, Lurz R, Breuer P, Schugardt N, Plassmann S, Xu KX, Warrick JM, Suopanki J, Wullner U, Frank R, Hartl UF, Bonini NM, Wanker EE. An arginine/lysine-rich motif is crucial for VCP/p97-mediated modulation of ataxin-3 fibrillogenesis. *EMBO J* 2006;25:1547–1558. [PubMed: 16525503]
31. Tam S, Geller R, Spiess C, Frydman J. The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. *Nat Cell Biol* 2006;8:1155–U211. [PubMed: 16980959]
32. Kitamura A, Kubota H, Pack CG, Matsumoto G, Hirayama S, Takahashi Y, Kimura H, Kinjo M, Morimoto RI, Nagata K. Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state. *Nat Cell Biol* 2006;8:1163–U224. [PubMed: 16980958]
33. Morimoto RI, Gidalevitz T, Ben-Zvi A, Ho K, Brignull H, Garcia S, Lima C, Fox S. Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Eur J Cell Biol* 2006;85:20–20.
34. Faber PW, Barnes GT, Srinidhi J, Chen JM, Gusella JF, MacDonald ME. Huntingtin interacts with a family of WW domain proteins. *Hum Mol Genet* 1998;7:1463–1474. [PubMed: 9700202]
35. Huang CC, Faber PW, Persichetti F, Mittal V, Vonsattel JP, MacDonald ME, Gusella JF. Amyloid formation by mutant huntingtin: Threshold, progressivity and recruitment of normal polyglutamine proteins. *Somat Cell Mol Genet* 1998;24:217–233. [PubMed: 10410676]
36. Lim J, Crespo-Barreto J, Jafar-Nejad P, Bowman AB, Richman R, Hill DE, Orr HT, Zoghbi HY. Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. *Nature* 2008;452:713–U1. [PubMed: 18337722]
37. Ross CA, Poirier MA. What is the role of protein aggregation in neurodegeneration? *Nat Rev Mol Cell Biol* 2005;6:891–898. [PubMed: 16167052]
38. Scherzinger E, Lurz R, Turmaine M, Mangiarini L, Hollenbach B, Hasenbank R, Bates GP, Davies SW, Lehrach H, Wanker EE. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* 1997;90:549–558. [PubMed: 9267034]

39. Perutz MF, Pope BJ, Owen D, Wanker EE, Scherzinger E. Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid beta-peptide of amyloid plaques. *Proc Natl Acad Sci U S A* 2002;99:5596–5600. [PubMed: 11960015]
40. Chen SM, Berthelie V, Hamilton JB, O’Nuallain B, Wetzel R. Amyloid-like features of polyglutamine aggregates and their assembly kinetics. *Biochemistry* 2002;41:7391–7399. [PubMed: 12044172]
41. Chen SM, Ferrone FA, Wetzel R. Huntington’s disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci U S A* 2002;99:11884–11889. [PubMed: 12186976]
42. Lee CC, Walters RH, Murphy RM. Reconsidering the mechanism of polyglutamine peptide aggregation. *Biochemistry* 2007;46:12810–12820. [PubMed: 17929830]
43. Ignatova Z, Gierasch LM. Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proc Natl Acad Sci U S A* 2006;103:13357–13361. [PubMed: 16899544]
44. Ignatova Z, Gierasch LM. Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein. *J Biol Chem* 2006;281:12959–12967. [PubMed: 16524881]
45. Scherzinger E, Sittler A, Schweiger K, Heiser V, Lurz R, Hasenbank R, Bates GP, Lehrach H, Wanker EE. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: Implications for Huntington’s disease pathology. *Proc Natl Acad Sci U S A* 1999;96:4604–4609. [PubMed: 10200309]
46. Altschuler EL, Hud NV, Mazrimas JA, Rupp B. Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides. *J Pept Res* 1997;50:73–75. [PubMed: 9273890]
47. Masino L, Kelly G, Leonard K, Trottier Y, Pastore A. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett* 2002;513:267–272. [PubMed: 11904162]
48. Sikorski P, Atkins E. New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils. *Biomacromolecules* 2005;6:425–432. [PubMed: 15638548]
49. Sharma D, Shinchuk LM, Inouye H, Wetzel R, Kirschner DA. Polyglutamine homopolymers having 8–45 residues form slablike beta-crystallite assemblies. *Proteins-Structure Function and Bioinformatics* 2005;61:398–411.
50. Perutz MF, Johnson T, Suzuki M, Finch JT. Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A* 1994;91:5355–5358. [PubMed: 8202492]
51. Perutz MF, Pope BJ, Owen D, Wanker EE, Scherzinger E. Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid β -peptide of amyloid plaques. *Proc Natl Acad Sci U S A* 2002;99:5596–5600. [PubMed: 11960015]
52. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc Natl Acad Sci U S A* 2006;103:16764–16769. [PubMed: 17075061]
53. Pappu RV, Wang X, Vitalis A, Crick SL. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch Biochem Biophys* 2008;469:132–141. [PubMed: 17931593]
54. Vitalis A, Wang XL, Pappu RV. Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories. *Biophys J* 2007;93:1923–1937. [PubMed: 17526581]
55. Wang XL, Vitalis A, Wyczalkowski MA, Pappu RV. Characterizing the conformational ensemble of monomeric polyglutamine. *Proteins-Structure Function and Bioinformatics* 2006;63:297–311.
56. Starikov EB, Lehrach H, Wanker EE. Folding of oligoglutamines: A theoretical approach based upon thermodynamics and molecular mechanics. *J Biomol Struct Dyn* 1999;17:409–427. [PubMed: 10636078]
57. Flory PJ. Thermodynamics of high polymer solutions. *The Journal of Chemical Physics* 1941;9:660–661.

58. Flory PJ. Thermodynamics of dilute solutions of high polymers. *The Journal of Chemical Physics* 1945;13:453–465.
59. Huggins ML. Solutions of long chain compounds. *The Journal of Chemical Physics* 1941;9:440.
60. Huggins ML. Some properties of solutions of long-chain compounds. *J Phys Chem* 1942;46:151–158.
61. Muthukumar M. Thermodynamics of polymer solutions. *The Journal of Chemical Physics* 1986;85:4722–4728.
62. Fields GB, Alonso DOV, Stigter D, Dill KA. Theory for the aggregation of proteins and copolymers. *J Phys Chem* 1992;96:3974–3981.
63. Grinberg NV, Dubovik AS, Grinberg VY, Kuznetsov DV, Makhaeva EE, Grosberg AY, Tanaka T. Studies of the thermal volume transition of poly(N-isopropylacrylamide) hydrogels by high-sensitivity differential scanning microcalorimetry. 1. Dynamic effects. *Macromolecules* 1999;32:1471–1475.
64. Chuang J, Grosberg AY, Tanaka T. Topological repulsion between polymer globules. *J Chem Phys* 2000;112:6434–6442.
65. Allegra G, Ganazzoli F, Raos G. The collapse of macromolecules in poor solvents. *Trends in Polymer Science* 1996;4:293–298.
66. Raos G, Allegra G. Chain collapse and phase separation in poor-solvent polymer solutions: A unified molecular description. *J Chem Phys* 1996;104:1626–1645.
67. Raos G, Allegra G. A cluster of chains can be smaller than a single chain: New interpretation of kinetics of collapse experiments. *Macromolecules* 1996;29:8565–8567.
68. Raos G, Allegra G. Macromolecular clusters in poor-solvent polymer solutions. *J Chem Phys* 1997;107:6479–6490.
69. Grosberg AY, Kuznetsov DV. Quantitative theory of the globule-to-coil transition. 1. Link density distribution in a globule and its radius of gyration. *Macromolecules* 1992;25:1970–1979.
70. Grosberg AY, Kuznetsov DV. Quantitative theory of the globule-to-coil transition. 2. Density correlation in a globule and the hydrodynamic radius of a macromolecule. *Macromolecules* 1992;25:1980–1990.
71. Grosberg AY, Kuznetsov DV. Quantitative theory of the globule-to-coil transition. 4. Comparison of theoretical results with experimental data. *Macromolecules* 1992;25:1996–2003.
72. Nishio I, Sun ST, Swislow G, Tanaka T. First observation of the coil-globule transition in a single polymer chain [6]. *Nature* 1979;281:208–209.
73. Nishio I, Swislow G, Sun ST, Tanaka T. Critical density fluctuations within a single polymer chain. *Nature* 1982;300:243–244.
74. Sun ST, Nishio I, Swislow G, Tanaka T. The coil-globule transition: Radius of gyration of polystyrene in cyclohexane. *The Journal of Chemical Physics* 1980;73:5971–5975.
75. Swislow G, Sun ST, Nishio I, Tanaka T. Coil-globule phase transition in a single polystyrene chain in cyclohexane. *Phys Rev Lett* 1980;44:796–798.
76. Oxtoby DW. Homogeneous nucleation: Theory and experiment. *J Phys: Condens Matter* 1992;4:7627–7650.
77. Ferrone F. Analysis of protein aggregation kinetics. *Methods Enzymol* 1999;309:256–274. [PubMed: 10507029]
78. Ferrone FA. Nucleation: The Connections Between Equilibrium and Kinetic Behavior. *Methods Enzymol* 2006;412:285–299. [PubMed: 17046664]
79. Vitalis, A.; Pappu, RV. ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem.* 2008. <http://10.1002/jcc.21005>
80. Imbert JB, Victor JM. Beyond Flory's theory: A computer aided phenomenology for polymers. *Molecular Simulation* 1996;16:399–419.
81. Imbert JB, Lesne A, Victor JM. Distribution of the order parameter of the coil-globule transition. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 1997;56:5630–5647.
82. Withers IM, Dobrynin AV, Berkowitz ML, Rubinstein M. Monte Carlo simulation of homopolymer chains. I. Second virial coefficient. *J Chem Phys* 2003;118:4721–4732.

83. Yang S, Levine H, Onuchic JN. Protein oligomerization through domain swapping: Role of inter-molecular interactions and protein concentration. *J Mol Biol* 2005;352:202–211. [PubMed: 16061250]
84. Milchev A, Paul W, Binder K. Off-lattice Monte Carlo simulation of dilute and concentrated polymer solutions under theta conditions. *The Journal of Chemical Physics* 1993;99:4786–4798.
85. Dobson CM, Swoboda BEP, Joniau M, Weissman C. The structural basis of protein folding and its links with human disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2001;356:133–145.
86. Talanquer V, Oxtoby DW. Crystal nucleation in the presence of a metastable critical point. *J Chem Phys* 1998;109:223–227.
87. Hoof RWW, Sander C, Scharf M, Vriend G. The PDBFINDER database: A summary of PDB, DSSP and HSSP information with added value. *Comput Appl Biosci* 1996;12:525–529. [PubMed: 9021272]
88. Richards FM, Kundrot CE. Identification of structural motifs from protein coordinate data: Secondary structure and first-level supersecondary structure. *Proteins: Structure, Function and Genetics* 1988;3:71–84.
89. Krishnan R, Lindquist SL. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* 2005;435:765–772. [PubMed: 15944694]
90. Derkatch IL, Uptain SM, Outeiro TF, Krishnan R, Lindquist SL, Liebman SW. Effects of Q/N-rich, polyQ, and non-polyQ amyloids on the de novo formation of the [PSI⁺] prion in yeast and aggregation of Sup35 in vitro. *Proc Natl Acad Sci U S A* 2004;101:12934–12939. [PubMed: 15326312]
91. Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc Natl Acad Sci U S A* 2007;104:2649–2654. [PubMed: 17299036]
92. Enderlein J. Collapsed but not folded: Looking with advanced optical spectroscopy at protein folding. *Chemphyschem* 2007;8:1607–1609. [PubMed: 17577904]
93. Bernstein SL, Wytenbach T, Baumketner A, Shea JE, Bitan G, Teplow DB, Bowers MT. Amyloid γ -protein: Monomer structure and early aggregation states of A γ 42 and its Pro19 alloform. *J Am Chem Soc* 2005;127:2075–2084. [PubMed: 15713083]
94. Teplow DB, Lazo ND, Bitan G, Bernstein S, Wytenbach T, Bowers MT, Baumketner A, Shea JE, Urbanc B, Cruz L, Borreguero J, Stanley HE. Elucidating amyloid γ -protein folding and assembly: A multidisciplinary approach. *Acc Chem Res* 2006;39:635–645. [PubMed: 16981680]
95. Wei G, Shea JE. Effects of solvent on the structure of the alzheimer amyloid- γ (25–35) peptide. *Biophys J* 2006;91:1638–1647. [PubMed: 16766615]
96. Bellesia G, Shea JE. Self-assembly of γ -sheet forming peptides into chiral fibrillar aggregates. *J Chem Phys* 2007;126.
97. Tarus B, Straub JE, Thirumalai D. Probing the initial stage of aggregation of the A γ 10–35- protein: Assessing the propensity for peptide dimerization. *J Mol Biol* 2005;345:1141–1156. [PubMed: 15644211]
98. Tarus B, Straub JE, Thirumalai D. Structures and Free-Energy Landscapes of the Wild Type and Mutants of the A γ 21–30 Peptide Are Determined by an Interplay between Intrapeptide Electrostatic and Hydrophobic Interactions. *J Mol Biol* 2008;379:815–829. [PubMed: 18479708]
99. Caflisch A. Computational models for the prediction of polypeptide aggregation propensity. *Curr Opin Chem Biol* 2006;10:437–444. [PubMed: 16880001]
100. Pellarin R, Caflisch A. Interpreting the Aggregation Kinetics of Amyloid Peptides. *J Mol Biol* 2006;360:882–892. [PubMed: 16797587]
101. Pellarin R, Guarnera E, Caflisch A. Pathways and Intermediates of Amyloid Fibril Formation. *J Mol Biol* 2007;374:917–924. [PubMed: 18028943]
102. Marchut AJ, Hall CK. Spontaneous formation of annular structures observed in molecular dynamics simulations of polyglutamine peptides. *Comput Biol Chem* 2006;30:215–218. [PubMed: 16678490]
103. Marchut AJ, Hall CK. Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations. *Biophys J* 2006;90:4574–4584. [PubMed: 16565057]

104. Tsukamoto K, Shimizu H, Ishida T, Akiyama Y, Nukina N. Aggregation mechanism of polyglutamine diseases revealed using quantum chemical calculations, fragment molecular orbital calculations, molecular dynamics simulations, and binding free energy calculations. *Journal of Molecular Structure: THEOCHEM* 2006;778:85–95.
105. Marchut AJ, Hall CK. Effects of chain length on the aggregation of model polyglutamine peptides: Molecular dynamics simulations. *Proteins-Structure Function and Bioinformatics* 2007;66:96–109.
106. Esposito L, Paladino A, Pedone C, Vitagliano L. Insights into structure, stability, and toxicity of monomeric and aggregated polyglutamine models from molecular dynamics simulations. *Biophys J* 2008;94:4031–4040. [PubMed: 18234827]
107. Ogawa H, Nakano M, Watanabe H, Starikov EB, Rothstein SM, Tanaka S. Molecular dynamics simulation study on the structural stabilities of polyglutamine peptides. *Comput Biol Chem* 2008;32:102–110. [PubMed: 18243803]
108. Barton S, Jacak R, Khare SD, Ding F, Dokholyan NV. The length dependence of the PolyQ-mediated protein aggregation. *J Biol Chem* 2007;282:25487–25492. [PubMed: 17591778]
109. Khare SD, Ding F, Gwanmesia KN, Dokholyan NV. Molecular origin of polyglutamine aggregation in neurodegenerative diseases. *Plos Computational Biology* 2005;1:230–235. [PubMed: 16158094]
110. Bennett MJ, Huey-Tubman KE, Herr AB, West AP, Ross SA, Bjorkman PJ. A linear lattice model for polyglutamine in CAG-expansion diseases. *Proc Natl Acad Sci U S A* 2002;99:11634–11639. [PubMed: 12193654]
111. Ignatova Z, Thakur AK, Wetzel R, Gierasch LM. In-cell aggregation of a polyglutamine-containing chimera is a multistep process initiated by the flanking sequence. *J Biol Chem* 2007;282:36736–36743. [PubMed: 17942400]
112. Mukai H, Isagawa T, Goyama E, Tanaka S, Bence NF, Tamura A, Ono Y, Kopito RR. Formation of morphologically similar globular aggregates from diverse aggregation-prone proteins in mammalian cells. *Proc Natl Acad Sci U S A* 2005;102:10887–10892. [PubMed: 16040812]
113. Poirier MA, Li HL, Macosko J, Cai SW, Amzel M, Ross CA. Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization. *J Biol Chem* 2002;277:41032–41037. [PubMed: 12171927]
114. Cecchini M, Rao F, Seeber M, Caflisch A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *The Journal of Chemical Physics* 2004;121:10748–10756. [PubMed: 15549960]
115. Smith AM, Jahn TR, Ashcroft AE, Radford SE. Direct Observation of Oligomeric Species formed in the Early Stages of Amyloid Fibril Formation using Electrospray Ionisation Mass Spectrometry. *J Mol Biol* 2006;364:9–19. [PubMed: 17005201]
116. Podlisny MB, Walsh DM, Amarante P, Ostaszewski BL, Stimson ER, Maggio JE, Teplow DB, Selkoe DJ. Oligomerization of endogenous and synthetic amyloid β -protein at nanomolar levels in cell culture and stabilization of monomer by Congo red. *Biochemistry* 1998;37:3602–3611. [PubMed: 9521679]
117. Andrews JM, Roberts CJ. A Lumry-Eyring nucleated polymerization model of protein aggregation kinetics: 1. Aggregation with pre-equilibrated unfolding. *J Phys Chem B* 2007;111:7897–7913. [PubMed: 17571872]
118. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 1999;314:141–151.
119. Sugita Y, Okamoto Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem Phys Lett* 2000;329:261–270.
120. Nymeyer H, Gnanakaran S, Garcia AE. Atomic Simulations of Protein Folding, Using the Replica Exchange Algorithm. *Methods Enzymol* 2004;383:119–149. [PubMed: 15063649]
121. Wyczalkowski MA, Pappu RV. Satisfying the fluctuation theorem in free-energy calculations with Hamiltonian replica exchange. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 2008;77.
122. Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins: Structure, Function and Genetics* 1999;35:133–152.

Figure 1. Coil-to-globule transitions for monomeric polyglutamine molecules of different chain lengths plotted as the variation of normalized chain size ζ (ordinate) with temperature (abscissa).

Figure 2.

Scaling of average internal distances $\langle R_{ij} \rangle$ between residues i and j as a function of sequence separation $|j-i|$ for monomeric polyglutamine. In each panel, the dashed curve plots $\langle R_{ij} \rangle$ as $R_0 |j-i|^{0.5}$, where $R_0 = 5.7 \text{ \AA}$, which is the expected profile at T_0 . By comparing data from simulations to the dashed curves we note that $T = 410 \text{ K}$ is a reasonable estimate for T_0 for polyglutamine modeled using the ABSINTH force field.

Figure 3. Cumulative distribution functions measuring the probability of sampling specific intermolecular distances between pairs of polyglutamine molecules.

Figure 4. Variation of the intermolecular excess pair interaction coefficients $B_{22}(T)$ as a function of temperature (top panel) and chain length, N (bottom panel).

Figure 5.

Plots showing correlations between temperature dependencies of $B_{22}(T)$ and chain density,

$\rho = \frac{N}{6\langle R_g \rangle^3}$ (units of $\text{\AA}^{-3}$) in panel (a) and normalized radius of gyration, ζ (units of $\text{\AA}$) in panel (b). Data are shown for Q_{30} and Q_{45} in the temperature range $298 \text{ K} \leq T \leq 360 \text{ K}$. This temperature regime corresponds to poor solvent conditions for polyglutamine in the ABSINTH model.

Figure 6.

Temperature dependencies of the fractional α -helical (f_α) and β -sheet (f_β) contents for Q₅, Q₁₅, Q₃₀, and Q₄₅, respectively. The squares (blue and black) quantify f_α and f_β using data from simulations of monomeric polyglutamine. Conversely, the diamonds (green and red) quantify f_α and f_β using data from simulations of pairs of polyglutamine molecules. Data are shown over a range of temperatures spanning the collapse regime and the transition regime up to the theta point.

Figure 7.

Temperature dependence of the variance in the number of intramolecular contacts for monomeric polyglutamine of different chain lengths – panel (a) – and for two globular proteins with well-defined folds *viz.*, the B1 domain of protein G (GB1) and engrailed homeodomain (ENH) – panel (b).

Figure 8. Cumulative distribution functions $F(R)$ quantifying the probability of realizing an intermolecular separation that is less than or equal to R for pairs of Q_{30} molecules that are either fully flexible (solid curves) or rigid (dashed curves) globules. Data were gathered from Metropolis Monte Carlo simulations that were performed with the simulation temperature $T=298K$.

Figure 9.

Schematic for the formation of higher order aggregates given a prescribed poorness of solvent. Our simulations have addressed steps (a), (b), and (c), albeit at very low copy numbers and high effective concentrations. In step (a) we find that monomer conformational equilibria favor non-specific collapsed states as opposed to a single folded species or an ensemble of extended conformations. Step (b) shows the favorable, spontaneous homodimerization and formation of disordered dimers. Step (c) shows that in our simulations disordered dimers are favored over ordered dimers. We have not probed step (d) in our simulations as this requires probing the formation of high energy species along a well-defined reaction coordinate. The schematic addresses possible routes to the ordered aggregate shown in the bottom right corner of the picture. Scenario I calls for the conformational reorganization within a large, disordered linear/spherical aggregate. This scenario resembles the concept of gelation or ordered aggregation, which requires a critical number of intermolecular contacts. Scenario II is that of Chen et al., which might be on-pathway and will require back tracking from the formation of non-specific oligomers/large linear aggregates. Scenario III depicts the formation of ordered aggregates via entanglement of swollen conformations that are also high energy species in a poor solvent. This scenario will also require back tracking of some sort in order to avoid the formation of non-specific spherical/linear aggregations depicted in the first half of scenario I.

Table 1

Overview of the magnitude of MMC sampling used for polyglutamine.

Simulation system	Number of independent replica exchange simulations	Simulation temperature of different replicas in K	Total number of MMC moves that make up a production run	No. of moves between replica exchange swaps
Q ₅ Monomer	4	298, 305, 315, 325, 335, 345, 355, 360, 370, 380, 390, 400, 410	2×10^7	0.5×10^5
Q ₁₅ Monomer			10^7	0.5×10^5
Q ₃₀ Monomer			2×10^7	1×10^5
Q ₄₅ Monomer			3×10^7	1×10^5
Q ₅ Dimer			2×10^7	1×10^5
Q ₁₅ Dimer			2×10^7	1×10^5
Q ₃₀ Dimer			4×10^7	1×10^5
Q ₄₅ Dimer			6×10^7	1.5×10^5
Q ₁₅ Monomer (High-T)	1	430, 450, 470, 490, 510, 530, 550	10^7	0.5×10^5
Q ₃₀ Monomer (High-T)		430, 450, 470, 500, 550, 600, 650, 700	2×10^7	10^5
Q ₄₅ Monomer (High-T)		420, 430, 450, 470, 500, 550, 600, 650, 700	3×10^7	4×10^5