Thermodynamics of β-Sheet Formation in Polyglutamine

Andreas Vitalis,[†‡] Nicholas Lyle,[†§] and Rohit V. Pappu[¶†‡§*]
[†]Center for Computational Biology, [‡]Molecular Biophysics Program, [§]Computational Biology Program, and [¶]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri

ABSTRACT The role of β-sheets in the early stages of protein aggregation, specifically amyloid formation, remains unclear. Interpretations of kinetic data have led to a specific model for the role of β-sheets in polyglutamine aggregation. According to this model, monomeric polyglutamine, which is intrinsically disordered, goes through a rare conversion into an ordered, metastable, β-sheeted state that nucleates aggregation. It has also been proposed that the probability of forming the critical nucleus, a specific β-sheet conformation for the monomer, increases with increasing chain length. Here, we test this model using molecular simulations. We quantified free energy profiles in terms of β-content for monomeric polyglutamine as a function of chain length. In accord with estimates from experimental data, the free energy penalties for forming β-rich states are in the 10–20 kcal/mol range. However, the length dependence of these free energy penalties does not mirror interpretations of kinetic data. In addition, although homodimerization of disordered molecules is spontaneous, the imposition of conformational restraints on polyglutamine molecules does not enhance the spontaneity of intermolecular associations. Our data lead to the proposal that β-sheet formation is an attribute of peptide-rich phases such as high molecular weight aggregates rather than monomers or oligomers.

INTRODUCTION

Nine neurodegenerative diseases are associated with polyglutamine expansion mutations (1). Ages of onset and severity of disease at onset are inversely correlated with the lengths of polyglutamine expansions (2). Proteolysis of mutant proteins releases fragments rich in polyglutamine that aggregate to form insoluble neuronal intranuclear inclusions (3). Aggregates are fibrillar and amyloidlike (4) with high β-sheet contents. Based on fiber diffraction data, Perutz et al. proposed a model in which the individual polyglutamine molecules within aggregates are arranged in a β-helical conformation (5). The fiber diffraction data have also been shown to be consistent with flat β-sheet architectures for individual polyglutamine peptides (6).

In contrast to structures adopted by individual polyglutamine molecules in fibrillar aggregates, monomeric polyglutamine molecules have spectroscopic signals that indicate a lack of well-defined secondary structures (7,8). It has also been shown that monomeric polyglutamine forms collapsed structures in water (9), and this observation is consistent with water being a poor solvent for polyglutamine (10). Monomeric polyglutamine is intrinsically disordered, irrespective of chain length, and these molecules sample heterogeneous ensembles of collapsed structures in water.

Chen et al. (11) proposed a model to connect intrinsically disordered monomers and β-sheet structures that are prominent in fibrillar forms. Analysis of kinetic data using a homogeneous nucleation model yielded a nucleus size of one (11). The results were interpreted as follows: Monomeric polyglutamine is in a preequilibrium between a folded, toxic, β-sheet structure and the disordered ensemble. Whenever the toxic fold is adopted, the thermodynamically unfavorable nucleus is populated and the resulting ordered conformation is elongated through monomer addition. A schematic of this proposal is shown in Fig. 1. The proposal put forth by Chen et al. is difficult to test experimentally because the critical nucleus is, by definition, a rare species.

Recently (12), we corroborated experimental results that polyglutamine chains spanning the pathological threshold range for Huntington's disease ($N \approx 37$) adopt disordered, compact conformations at the monomer level. We also showed that they associate spontaneously to form disordered dimers. These studies were carried out using the ABSINTH implicit solvation model combined with the all atom OPLS-AA/L force field (12,13). Here we use the same combination of force field and implicit solvation model to answer a set of questions and test all aspects of the homogeneous nucleation model proposed by the analysis of Chen et al. (11) and Bhattacharya et al. (14). The questions of interest are as follows:

- What is the probability that monomeric polyglutamine forms structures with high β-content under ambient conditions?
- Are such species metastable states along a reaction coordinate that measures the net β-content of the chain?
- How does the likelihood of forming species high in β-content vary with chain length?
- If we bias individual polyglutamine chains toward structures with high β-content, do we observe a change in the spontaneity of dimer formation?
- Finally, do our results support the idea of a structure-driven, homogeneous nucleation pathway for polyglutamine aggregation?

Submitted February 23, 2009, and accepted for publication May 1, 2009.

Andreas Vitalis and Nicholas Lyle contributed equally to this work.

*Correspondence: pappu@wustl.edu

Editor: Ruth Nussinov.

© 2009 by the Biophysical Society

0006-3495/09/07/0303/9 $2.00

doi: 10.1016/j.bpj.2009.05.003

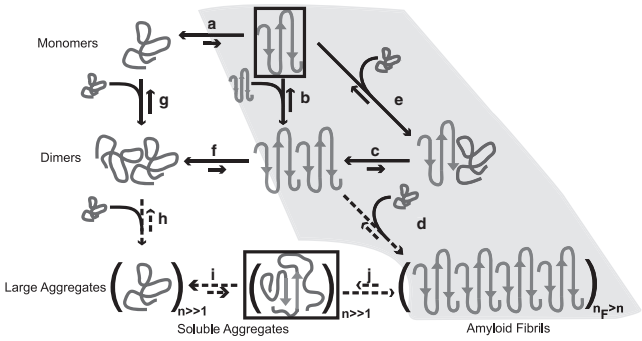

FIGURE 1 Schematic of possible aggregation pathways for polyglutamine in vitro; n denotes the number of polyglutamine molecules within a disordered aggregate and n_F denotes the number of polyglutamine molecules within an ordered amyloid fibril. The ordered amyloid fibril rich in β-sheets is shown in the bottom-right corner of the schematic. The gray-shaded region encompasses steps *a*, *e*, *c*, and *d* and depicts the homogeneous nucleation proposal of Chen et al. (11). We investigated the thermodynamics of step *a*, which indicates that the formation of ordered conformations is thermodynamically unfavorable. Monomeric polyglutamine prefers disordered, collapsed conformations, to the left of step *a*. Step *b* pertains to the thermodynamics of interactions between chains that have been restrained to adopt ordered conformations. Associativities of restrained chains—step *b*—are akin to the associativities of unrestrained chains—step *g*. However, the likelihood that chains will sample the associations shown in step *b* is very small, because this is tied to the equilibria in step *a*, which requires the population of the conformations with high β-content. Fig. 3 shows that this is highly unlikely. Similarly, step *f* shows that disordered dimers are thermodynamically favored to dimers with high β-contents in individual chains. This is the result of linkage to step *a* as discussed above. The aggregates achieved in step *h* are likely to be large (in terms of *n*) and exhibit spherical, "liquid-like" (31,32,34) organization of polyglutamine chains around each other. Step *i* depicts a slow conformational conversion of individual/small numbers of chains to β-sheets. This slow step is likely to lead to the creation of an ordered template for fibril formation via monomer or oligomer addition and elongation to yield the ordered amyloid fibril. Steps *a*, *b*, and *g* are anchored in the collection of data generated in this work and previous studies. However, the reversible associations depicted in step *h* and the conformational conversions depicted in step *i* are yet to be tested.

The rest of this article is organized as follows: First, we introduce the details of our methodology. This is followed by a presentation of results that allow us to answer the questions raised above and we conclude with a summary and discussion of our results.

MATERIALS AND METHODS

Degrees of freedom and force field

We simulated one- or two-polyglutamine molecule *n*-acetyl-$(Gln)_N$-*n*′-methylamide; the chains were modeled in atomic detail and the number of glutamine residues was set to $N = 5$, 15, 30, and 45, respectively. Chains with N glutamine residues are denoted as Q_N. Markov chain Metropolis Monte Carlo (MC) Simulations were performed in the canonical ensemble and molecules were enclosed in a spherical droplet of radius 200 Å, which was enforced using a stiff harmonic boundary potential. The degrees of freedom were the backbone ϕ, ψ, ω, and side-chain χ-dihedral angles. For MC simulations with two chains, rigid-body coordinates, namely center-of-mass translations and rotations, were included as additional degrees of freedom. Bond lengths and bond angles were held fixed at values prescribed by Engh and Huber (15).

We used parameters from the OPLS-AA/L force field (16) with appropriate modifications and the ABSINTH implicit solvent model (13). The energy function is

$$E_{\text{total}} = W_{\text{solv}} + W_{\text{el}} + U_{\text{LJ}} + U_{\text{tor}}. \quad (1)$$

Here, W_{solv} is the direct mean-field interaction term that captures the transfer of a polypeptide solute in a specific conformation from the gas phase into the continuum solvent. W_{el} denotes the mean-field electrostatic term; the dielectric constant of the solvent is set to be $\varepsilon = 78$. U_{LJ} models van der Waals interactions using the Lennard-Jones (LJ) model. Parameters for the LJ hard-sphere radii and well depths are based on heats of fusion data for model compounds and are different from the choices made in standard force fields. This choice allows us to omit many of the torsional potentials because the rotational barriers and minima are captured by excluded volume interactions alone (13). U_{tor} represents torsional terms applied to dihedral angles that cannot be captured by U_{LJ}. For polyglutamine, we use torsional potentials taken from OPLS-AA/L to maintain peptide dihedral angles (ω) in predominantly *trans*-configurations.

Salient features of the interplay between the W_{solv} and W_{el} terms are summarized as follows: Polypeptide chains are decomposed into a set of distinct solvation groups; for capped polyglutamine, the solvation groups are $N+1$ backbone secondary amides and N side-chain primary amides. W_{solv} is a sum of contributions from each solvation group and for each of these groups we use experimentally measured free energies of solvation of appropriate model compound analogs as references for fully solvated states. The degree of solvent accessibility modulates the direct mean-field interaction and consequently W_{solv} varies with conformation. Solvent-accessible volume fractions are used as the metric for solvent accessibility and this is used to evaluate the solvation states υ_k^i for atom k in solvation group i. Electrostatic interactions between nonbonded solute atoms with partial charges are fully screened by the continuum dielectric if the atoms are fully exposed to solvent. The screening of polar interactions is conformation-dependent and the solvation states of atoms υ_k^i determine the extent to which the screening of electrostatic interactions is modulated. ABSINTH is built upon the distinct strengths of the generalized Born (GB) (17–19) and EEF1 (20) models. In accord with EEF1, the process of transferring solutes from the gas phase into the continuum solvent is treated in "one shot" without attempting to parse the distinct contributions from the polar and nonpolar components. However, ABSINTH deviates from EEF1 in the way electrostatic interactions between solute atoms are handled. No explicit or implicit distance-dependence is assumed for the dielectric response. Instead, solvation states of individual atoms υ_k^i are used to determine the extent to which the screening of electrostatic interactions is to be modulated by the protein environment. ABSINTH therefore captures the main strengths of the GB model while retaining the efficiency of the EEF1 model and it may be viewed as an effective interpolation between GB and EEF1.

Sampling methodology

Move sets for MC simulations included pivots about random backbone torsions, perturbations about random dihedrals, randomized side-chain rotations, and randomized changes to rigid body coordinates. Details regarding the move sets are summarized in Table S1 in Supporting Material. One of our goals was to quantify free energy penalties associated with sampling of β-rich conformations for monomeric polyglutamine. We defined a reaction coordinate f_β to assess global β-content within a molecule and used umbrella sampling to enhance the sampling of the low-likelihood regions of conformational space. Results of independent umbrella sampling calculations were stitched together using weighted histogram analysis methods (WHAM) (21,22) to generate unbiased potentials of mean force (PMFs) as a function of f_β, which is defined as

$$f_\beta = \frac{1}{N}\sum_{i=1}^{N} f_\beta^{(i)},$$

where

$$f_\beta^{(i)} = \begin{cases} 1.0 & \text{if } (\phi_i, \psi_i) \text{ belongs to the } \beta\text{-basin} \\ \exp\left(-\tau_\beta d_{(i)}^2\right) & \text{otherwise} \end{cases}, \quad (2)$$

$$d_{(i)}^2 = \left\{ \left(\sqrt{\left[(\phi_i - \phi_\beta)\bmod 2\pi\right]^2 + \left[(\psi_i - \psi_\beta)\bmod 2\pi\right]^2} - r_\beta \right) \bmod 2\pi \right\}^2. \quad (3)$$

Here, N is the number of glutamine residues in the sequence, mod2π terms correct for periodicity effects associated with distance calculations in angular space; (ϕ_β,ψ_β) define the reference (ϕ,ψ) values for an individual residue. If residue i adopts ϕ- and ψ-angles that lie within a circle of radius r_β, then the parameter $f_\beta^{(i)}$ is set to unity; otherwise, $f_\beta^{(i)}$ has a value between 0 and 1; the precise value is determined by two parameters, namely $d_{(i)}^2$ and τ_β. The latter is the width of the Gaussian function and ensures a continuous function. We used: $(\phi_\beta,\psi_\beta) = (-152^\circ, 142^\circ)$, $r_\beta = 50^\circ$, and $\tau_\beta = 0.002$ deg^{-2}.

We assessed the validity of f_β as a measure of β-content by quantifying its ability to estimate β-content in proteins of known three-dimensional structures. We used PDBSelect (23) to create a database of 3693 nonredundant protein structures from the Protein Data Bank. Sequences in this dataset have <25% sequence identity with each other. For each structure in the dataset, we calculated the f_β values and their DSSP-E score (24), normalized by the number of residues, as an alternative to measure the degree of ordered β-sheet. Fig. 2 shows the correlation between f_β and fractional DSSP-E scores. The correlation coefficient is 0.83. Fig. S1 shows ribbon drawings of three-dimensional structures for five structures from the database. Since we do not have definitive prior knowledge of the type of ordered β-sheets that polyglutamine molecules adopt in fibrillar aggregates, we used f_β instead of fractional DSSP-E scores as a reaction coordinate for assessing the bias toward structures with high or low β-contents.

Simulations for monomeric polyglutamine were carried out at 298 K. For each chain length, we performed 17 sets of distinct umbrella sampling simulations and in each simulation, f_β was restrained to one of 17 target f_β^0 values: [0.0, 0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.75, 0.8, 0.9, 0.95, 1.0]. Each simulation was biased toward a target value f_β^0 using a harmonic restraint potential $U_{restr} = k(f_\beta - f_\beta^0)^2$ that was added to the energy function shown in Eq. 1. The number of (ϕ,ψ) pairs that contribute to f_β increases with N and the value of k varied with N. We used values of k = 25 kcal/mol and 75 kcal/mol for Q_5 and Q_{15}, and k = 150 kcal/mol for Q_{30} and Q_{45}, respectively. Therefore, k varies from 1.7 kcal/mol (Q_{45}) to 2.5 kcal/mol (Q_5, Q_{15}, and Q_{30}) per restrained degree of freedom.

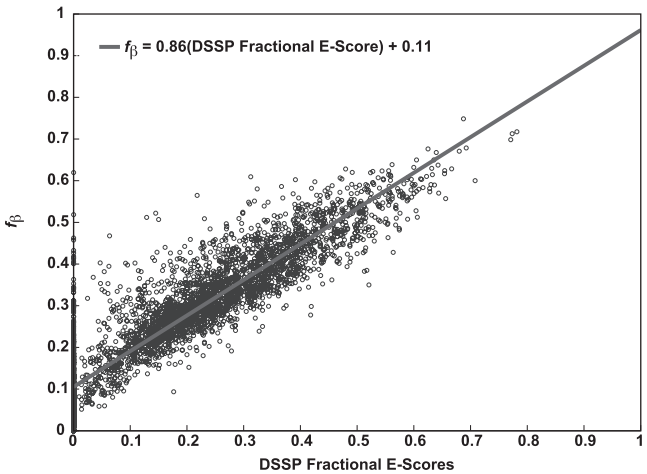

FIGURE 2 Correlation between fractional DSSP-E scores and f_β: The solid line is the line of best fit that quantifies the strength and direction of the linear correlation between f_β values and fractional DSSP-E scores. Parameters for the slope and intercept are shown in the inset. Structures that have high fractional DSSP-E scores also have high f_β values, although there is some scatter about the line of best fit. For ~27% of the structures in the dataset, the fractional DSSP-E scores are zero. Although the f_β values for most of these structures are small ($\leq$0.3), they span a finite range of f_β values.

For each window, the starting conformation was extracted at random from an ensemble of self-avoiding random walks. For Q_5, Q_{15}, and Q_{30} the first 10^6 MC steps were used for equilibration followed by 4×10^7 steps of production. For Q_{45}, we used 1.5×10^6 steps of equilibration and 6×10^7 steps of production. Sampling was enhanced using replica exchange MC in f_β-space (25). For each chain length, we performed three independent replica exchange umbrella sampling MC runs. The quality of sampling was assessed by computing statistics for the extent of overlap of f_β histograms between adjacent windows and statistics for replica exchange. These details are documented in Fig. S3, Fig. S4, and Fig. S5.

The WHAM protocol has two components, namely iterative computation of relative free energies between adjacent windows and the construction of unbiased PMFs. Data from restrained simulations were analyzed using WHAM in three different ways. In the first two methods, both components of the WHAM protocol are used, with the only difference being the method used to compute error bars for the unbiased PMFs. The first method reports error bars as standard errors and the second method uses a bootstrap procedure to estimate error bars. For the third method, the relative free energies between adjacent windows were computed using thermodynamic integration (TI) and the unbiased PMF was computed using WHAM. All the details regarding error analysis and TI-WHAM are provided in the Supporting Material.

For polyglutamine dimer simulations, we combined MC simulations with thermal replica exchange. In two of the three sets of simulations, each chain was restrained to target f_β^0 values of = 0.75 and 1.0, respectively, whereas the third simulation set involved unrestrained molecules. For each chain length, we carried out three independent replica exchange runs. The following Kelvin temperature schedule was used for the replica exchange simulations: [298, 305, 315, 325, 335, 345, 355, 360, 370, 380, 390]. The temperature schedule was based on data for coil-to-globule transitions of unrestrained monomeric polyglutamine. As shown in Fig. S2, temperature modulates the solvent quality. Temperatures in the range $T \leq 355$ K correspond to the poor solvent regime, where monomeric polyglutamine prefers globules; temperatures in the range 355 K $< T <$ 420 K are in the transition region between globule and coil, and higher temperatures correspond to swollen coils. The θ-point (T_θ) was found to be ~390 K (see caption to Fig. S2). We wish to quantify the spontaneity of intermolecular associations in the poor solvent regime. The overlap between coil and globule ensembles is small and decreases with increasing N. Therefore, we set the upper limit for the replica exchange temperature schedule to be T_θ, to ensure that the replicas were used judiciously. We quantified overlap statistics and acceptance ratios for swaps between adjacent thermal replicas to demonstrate the adequacy of the protocol, and details are in Fig. S6, Fig. S7, and Fig. S8.

RESULTS

Conformations with high β-content are thermodynamically unfavorable for monomeric polyglutamine

Fig. 3 *A* shows free energy profiles for monomeric polyglutamine of four chain lengths. Fig. 3, *B* and *C*, shows the same free energy profiles, except that the profile in Fig. 3 *B* shows the bootstrap errors, whereas Fig. 3 *C* shows the free energy profiles constructed using the TI-WHAM procedure. The free energy profiles along f_β are smooth and possess broad

minima for $N \geq 15$. For Q_5, the PMF is essentially flat because the peptide is short and cannot collapse on itself. For longer chains, the minima are located at values of $f_\beta \approx 0.3$, which is the value preferred by unrestrained polyglutamine chains. The Q_{15} chain possesses a tendency for forming α-helical segments (12). This helix propensity is the reason for the flatness of the free energy profile for values of $f_\beta < 0.3$ for this peptide. Decreased α-helical propensities for longer chain lengths lead to an increase in the free energy penalty for $f_\beta < 0.3$.

The free energy profiles do not show evidence for distinct local minima. This is confirmed by analyzing the derivatives of each of the PMFs. In the harmonic limit with a single minimum, a mean force profile will be a straight line and this is what we observe as shown in Fig. 3 *D*. Therefore, monomeric polyglutamine does not access metastable conformations with high f_β values. Bhattacharya et al. (14) estimated the nucleation free energy for Q_{47} to be roughly 12 kcal/mol. Beta-rich structures will have f_β values between 0.6 and 0.7 (Fig. 2 and Fig. S1). Free energy penalties associated with accessing such structures are in the range of 10–15 kcal/mol for Q_{45} (Fig. 3 *A*). This is in the range of the estimate obtained by Bhattacharya et al.

Bhattacharya et al. also proposed that aggregation rates increase with increasing chain length because free energy penalties associated with forming β-rich structures decrease with increasing chain length (11,14). The calculated free energy profiles do not support this hypothesis. The free energy penalty for forming structures with high f_β values increases with increasing chain length. However, this increase in free energy penalty levels off for Q_{45} because the PMFs for Q_{30} and Q_{45} resemble each other—more so than those for Q_{30} and Q_{15}. We previously showed that the stability of nonspecifically collapsed states increases with increasing chain length (12). Here, we show that the likelihood of populating structures with high f_β values decreases with increasing chain length. The former observation is consistent with the physics of flexible polymers in a poor solvent, whereas the latter observation is congruent with the idea that energetic frustration (26) makes it difficult for flexible, homopolymeric polyglutamine to adopt well-ordered three-dimensional structures in their monomeric forms.

Restraining monomeric polyglutamine to high f_β values provides access to ordered β-sheet conformations

Fractional DSSP-E scores (24) based on analysis of hydrogen-bonding patterns are a complementary measure of secondary structure. Fig. 4 plots data for monomeric polyglutamine with fractional DSSP-E scores along the ordinate and f_β values along the abscissa. Data were extracted from the simulations that were performed in the presence of restraints on f_β. We do not observe any structures for which the f_β value is low when the fractional DSSP-E score is high. Structures with characteristic β-sheet hydrogen-bond patterns have high fractional DSSP-E scores as well as high f_β scores. However, for moderate to high f_β values, there is a substantial spread in the observed fractional DSSP-E scores, implying that polyglutamine samples disordered conformations lacking regular backbone-backbone hydrogen bonds even when f_β is restrained to a high value.

Persistence of disorder in the presence of restraints results from a diverse registry of intramolecular hydrogen bonds

We asked whether disordered structures contain large numbers of hydrogen bonds, which are either unsatisfied or

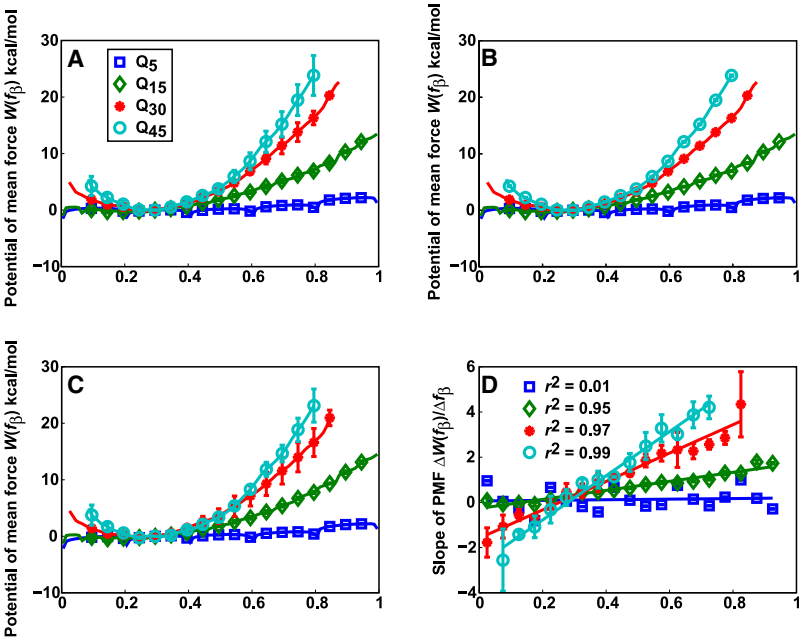

FIGURE 3 Potentials of mean force (PMFs) for monomeric polyglutamine chains of different lengths. The profiles are plotted as a function of f_β, which is a reaction coordinate that measures β-content. (*A*) PMFs with standard errors, (*B*) with bootstrap errors, and (*C*) PMFs that result from using TI-WHAM with a coarse f_β schedule. Details are described in the Supporting Material. (*D*) Lines of best fit to the data for the derivatives of the PMFs shown in panel *A*. The insets show the Pearson correlation coefficients—r^2—that diagnose the strength and direction of the hypothesis that the PMF derivatives (forces) are linear. The slopes and intercepts for the lines of best fit are as follows: Q_5 (0.25, −0.05), Q_{15} (2.00, −0.27), Q_{30} (5.74, −1.38), and Q_{45} (9.5, −2.65). Slopes have units of kcal/mol-f_β and intercepts have units of kcal/mol.

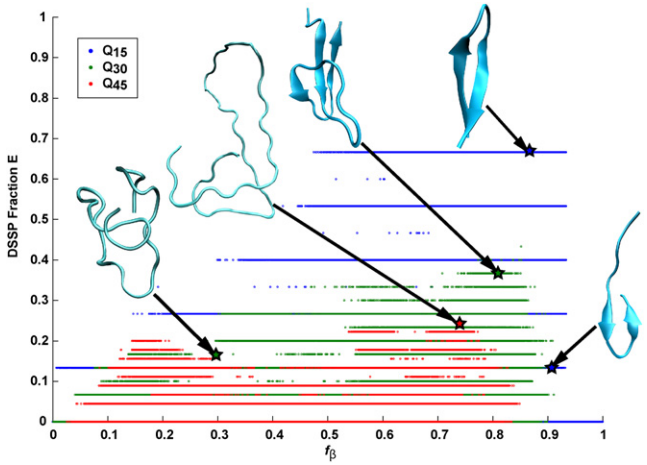

FIGURE 4 Scatter plot of all recorded snapshots in all simulations for Q_{15}, Q_{30}, and Q_{45} correlating the fractional β-content according to DSSP with the values for f_β at 298 K. Dots of different colors correspond to chains of different length. Representative points are marked using stars and the corresponding structures are shown in cartoon representation. Graphics were generated using VMD (35). Note that the fractional β-content according to DSSP is an inherently discrete quantity for chains of finite length. Q_5 is not shown since the chain is too short to have nonzero DSSP-E scores.

assumed to be satisfied by the solvent. In Fig. 5, we plot the average number of intramolecular hydrogen bonds per backbone and side-chain oxygen (acceptor) atom in monomeric polyglutamine as a function of chain length. All four types of hydrogen bonds including backbone acceptor to backbone donor, side-chain acceptor to side-chain donor, backbone acceptor to side-chain donor, and side-chain acceptor to backbone donor, make roughly equivalent contributions to the total hydrogen-bond inventory. No obvious preference for specific intramolecular contacts is seen, even in the presence of restraints. Summing up around the acceptor sites, the average number of hydrogen bonds per backbone side-chain acceptor is less than unity, indicating that a substantial fraction of hydrogen bonds is satisfied by the solvent. The average number of intramolecular hydrogen bonds per acceptor is generally larger for the unrestrained case. Restraints to high values of f_β increase the number of solvent-exposed acceptors leading to increased interfaces with the surrounding solvent. Despite the restraints, the peptides prefer the full spectrum of intramolecular and chain-solvent contacts.

Dimerization of polyglutamine remains spontaneous in the presence of restraints

We asked whether the spontaneity of homodimerization is altered when each chain is restrained to adopt high values of f_β. We simulated dimerization of polyglutamine for chains of length $N = 5$, 15, 30, and 45 as a function of temperature and two different values for the target restraint, $f_\beta{}^0$. If the homogeneous nucleation model applies for polyglutamine aggregation, then associations involving restrained chains should be stronger than the associations involving unrestrained chains. To quantify associativity, we computed a temperature-dependent excess interaction coefficient, $B_{22}(T)$, which we introduced in previous work and recapitulate below:

$$B_{22}(T) = \frac{\int_{R=0}^{R=d_{\text{droplet}}} [F_{T=T_\theta}(R) - F_T(R)]R^2 dR}{\int_{R=0}^{R=d_{\text{droplet}}} F_{T=T_\theta}(R)R^2 dR}. \quad (4)$$

Here, $F_T(R)$ is the cumulative distribution function for the intermolecular distance at temperature T; $d_{\text{droplet}} = 400$ Å is the diameter of the simulation system. When spontaneous associations are favored compared to T_θ, $B_{22}(T)$ is negative. If chains interact the way they would at T_θ, then $B_{22}(T)$ is zero.

Fig. 6 plots $B_{22}(T)$ as a function of T for different chain lengths in the presence and absence of restraints on f_β. The associativity of chains is length- and temperature-dependent, but it does not vary with the presence or absence of

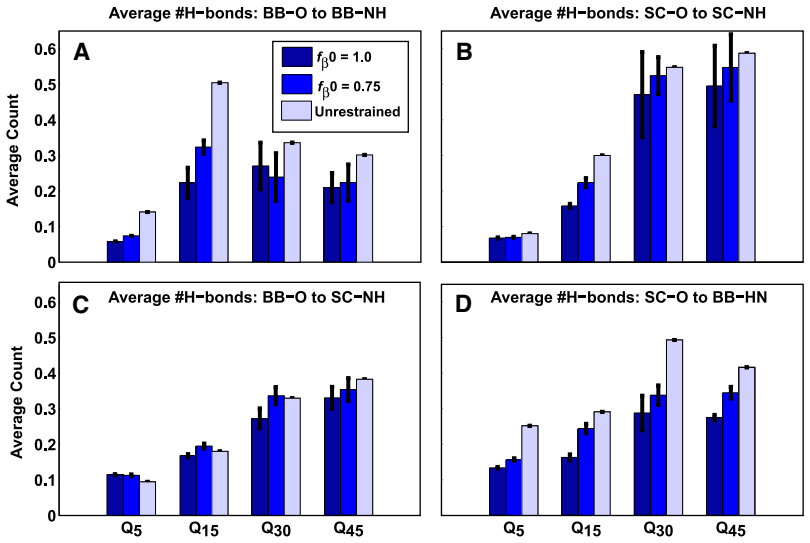

FIGURE 5 Average number of hydrogen bonds per acceptor oxygen atoms. Data are shown for hydrogen bonds around the backbone (*A* and *C*) and side-chain (*B* and *D*) acceptor oxygen atoms, respectively. *BB* denotes backbone and *SC* denotes side chains. Values are shown for $T = 298$ K and three different chain lengths (Q_{15}, Q_{30}, Q_{45}) and restraint values. Hydrogen bonds were determined using the general definition introduced by Kabsch and Sander (24).

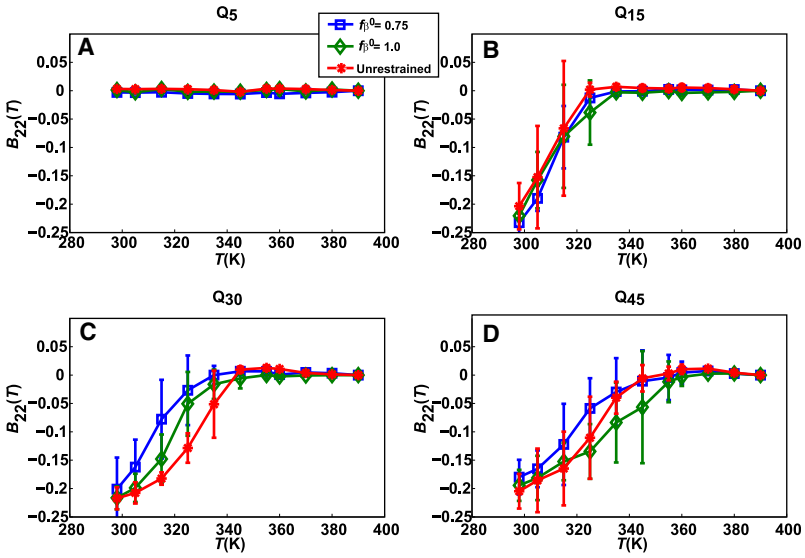

FIGURE 6 Plots of $B_{22}(T)$ as a function of temperature. Each panel shows $B_{22}(T)$ extracted from simulations with unrestrained chains and simulations where each chain in the simulation has a target restraint of $f_\beta^0 = 0.75$ or $f_\beta^0 = 1.0$.

conformational restraints. Fig. 6 *A* shows that the short Q_5 peptide remains nonassociative across the entire simulated temperature range. For Q_{15}, the differences in associativity with and without restraints are statistically insignificant. At the lowest temperatures, B_{22} is negative for restrained as well as unrestrained chains. It increases monotonically with increasing temperature and reaches zero between 320 and 340 K. This temperature range represents the stability limit for the associated dimer. Fig. 6, *B* and *C*, show the same data for Q_{30} and for Q_{45}. As chain lengths increase, the stability limit for associated dimers shifts to higher temperatures. Within the statistical accuracy of the data, chains with restraints are as associative (not more associative) as unrestrained chains.

For high values of f_β, the surface/volume ratio increases with respect to the unrestrained chains (see Fig. S9). This does not translate into increased associativity with respect to the unrestrained chains. Instead, the actual poorness of the solvent (the value of T) remains the main determinant of the spontaneity of intermolecular associations. This observation is confounding since solute-solvent interfaces increase for monomeric polyglutamine when restraints toward high f_β are imposed. We quantified the solvent-solute interface for monomeric polyglutamine by asking if interfacial energies were insensitive to the presence or absence of conformational restraints. We decomposed the system energies (excluding the restraint potential) into bulk (volumetric) and surface contributions (27) by fitting the system energies to the functional form shown in Eq. 5. Here, C_1 and C_2 measure the effective strengths of volumetric self-interactions and surface terms, respectively,

$$\frac{\langle U_{\text{total}} - U_{\text{restr}} \rangle}{N} = C_1\left(T, f_\beta^0\right) + C_2\left(T, f_\beta^0\right) \cdot N^{-1/3}, \quad (5)$$

where N is the chain length, U_{total} is the system energy, and U_{restr} is the restraint potential energy. Fig. 7 plots C_1 and C_2 as a function of temperature for different values of f_β^0. When compared to the unrestrained simulations, the C_1 term in the collapse regime is less favorable in the presence of restraints indicating the high free energy of structures with high values of f_β. Additionally, for a given temperature, values for C_2 are more positive because larger interfaces with the surrounding solvent are more unfavorable than the smaller solute-solvent interfaces made by unrestrained chains. As $T \rightarrow T_\theta$, C_2 approaches zero indicating that the interface is indifferent with respect to interactions with either solvent or the chain. This is Flory's definition of the θ-state. For temperatures $> T_\theta \approx 390$ K, C_1 converges to a value of ~–20 kcal/mol. This is the approximate mean-field energy for a fully

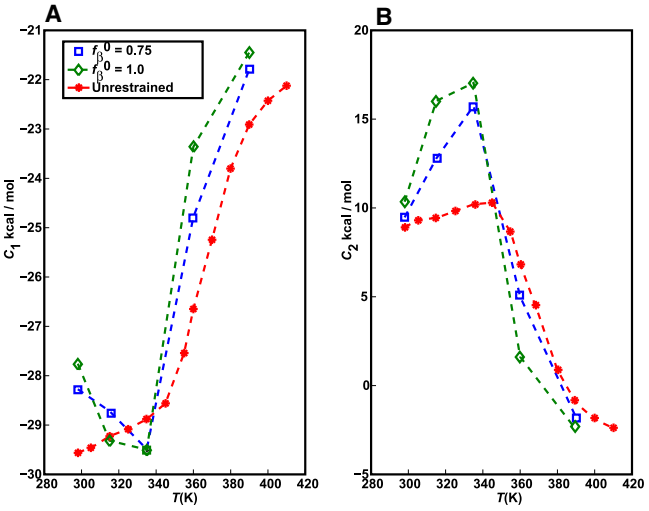

FIGURE 7 Energy density C_1 (*A*) and surface energy term C_2 (*B*) for monomeric polyglutamine. Data were obtained for unrestrained polyglutamine and two other simulations with restraints on f_β. The quality of the fits underlying these data cannot be accurately assessed since they are fits to data from only three chain lengths (because we exclude Q_5 excluded from analysis, since it is too short for a volumetric term to contribute).

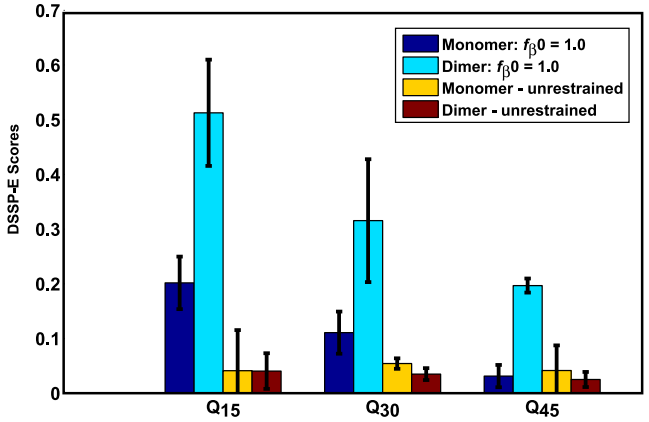

FIGURE 8 Bar plots comparing the average fractional DSSP-E scores in simulations of monomeric polyglutamine to simulations of dimeric polyglutamine at 298 K. Data are shown for three chain lengths using data from simulations with unrestrained chains as well as data from simulations for $f_\beta^0 = 1.0$.

solvated glutamine residue within the ABSINTH Hamiltonian (13), and is the expected limiting value for a chain preferring chain-solvent interactions to chain-chain interactions. Therefore, in the presence of restraints toward high target values of f_β, the chains swell and form energetically unfavorable interfaces with the surrounding solvent. However, these features do not translate into differences in B_{22} compared to the unrestrained chains.

Restraints toward high values of f_β promote the formation of canonical β-sheets

Fig. 8 shows the average fractional DSSP-E scores observed in monomer and dimer simulations with f_β^0=1 compared to the data obtained from simulations with unrestrained chains. When the free energy penalty for sampling high f_β values is prepaid, the resultant increase in solute-solvent interface at the monomer level results in increased β-sheet formation through a coupling between inter- and intramolecular interactions. This difference is suggestive of intermolecular interfaces being responsible for promoting β-sheet content. To provide a visual perspective of increased β-sheet content in dimers formed with restrained chains, Fig. S10 shows a scatter plot in the space of the two parameters, f_β and fractional DSSP-E scores. Fig. 9 shows a bar plot of intermolecular hydrogen bonds to quantify the interactions in intermolecular interfaces. Intermolecular hydrogen bonds predominantly involve glutamine side chains. There is a significant enhancement in intermolecular backbone-backbone hydrogen bonds for Q_{15} and Q_{30} when $f_\beta^0 = 1.0$ and is congruent with data in Fig. 8. Interestingly, the per-molecule PMF obtained via WHAM for two Q_{30} molecules in the simulation system is virtually identical to the one obtained for the monomer (data not shown). Hence, we conjecture that larger oligomers must form to facilitate spontaneous conversion of individual chains to β-rich conformations.

DISCUSSION

Monomeric polyglutamine encompassing the threshold length for Huntington's disease does not adopt conformations rich in β-content. Such conformations have been proposed as putative nuclei for polyglutamine aggregation (11). Our estimate of the free energy penalty associated with forming structures with high β-content is in the range of estimates proposed in the literature (14). However, these penalties increase with increasing chain length and may plateau past a certain chain length. Our observations contradict the expectations of Bhattacharya et al. (14), who proposed that the free energy penalties in question should decrease with increasing chain length if the stability of the monomeric nucleus is the source of the increased rate of aggregation with increasing chain length. In our calculations,

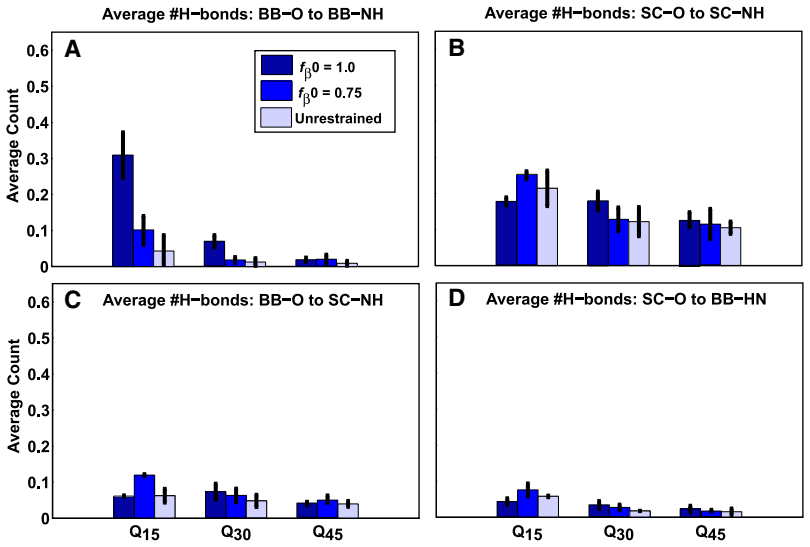

FIGURE 9 Average number of hydrogen bonds per acceptor oxygen atoms. This plot is similar to Fig. 5 except that the hydrogen-bond statistics are shown for simulations with two molecules. Only intermolecular hydrogen bonds are shown. Q_5 is excluded from this plot, since no intermolecular hydrogen bonds are detected in these simulations.

coil-to-globule transitions for monomeric polyglutamine (Fig. S2) as well as the spontaneities of intermolecular association show clear dependence on chain length. In a poor solvent, longer chains form disordered globules of increased stability. The more stable the globules, the more favorable the dimers (12), and the harder it is to dissociate the dimers. Although conformational restraints toward high f_β values lead to increases in solute-solvent interfaces and increased β-content, these do not translate into increased spontaneity for intermolecular associations. This observation points to poorness of the solvent, as opposed to the associativities ascribed to specific structures, as the invariant and generic driving force for promoting aggregation of homopolymers. Even though high β-content remains just as thermodynamically unfavorable at the dimer level as at the monomer level, the formation of intermolecular interfaces between restrained chains promotes the formation of canonical backbone-driven β-sheet structures. Other approaches to simulating polyglutamine conformational equilibria and aggregation have been published (28,29). Clearly, comparative calculations using different models are needed to test the model dependence of our predictions and this is part of ongoing work.

Recent experiments (30) suggest that β-sheet formation is a feature of large aggregates. It was also shown that large nonspecific aggregates form before nucleated formation of ordered fibrils. These observations are consistent with extrapolations from our calculations, specifically the data shown in Fig. 9. In Fig. 1, we summarize conceivable routes from the ensemble of disordered globules to the high molecular weight, ordered fibrillar aggregates. The homogeneous nucleation model, with a monomeric nucleus, is highlighted in gray. An alternative proposal, consistent with our results and recent experimental data of Lee et al. (30), is follows: Disordered globules reversibly associate to form a broad distribution of disordered oligomers. These oligomers can be described as being ''molten'' (31) or ''liquid-like'' (32) and may be large enough to be referred to as ''mesoglobules'' (33). Molten oligomers are peptide-rich microphases, and chains in the interior are solvated by other chains. These peptide-peptide interfaces should be more favorable than peptide-solvent interfaces in a poor solvent. Peptides within concentrated droplets can become indifferent to preferring intramolecular interactions to intermolecular interactions and consequently, individual chains can expand. In the ensuing conformational sampling, a presumed slow step, chains can converge on energetically favorable conformations that are high in β-content. Our proposal is similar to those of others for polyglutamine (30), Aβ (34), and the N-domain Sup35 (31) but needs to be tested using appropriate simulations.

SUPPORTING MATERIAL

Ten figures and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)00906-0.

This work was supported by grant No. 5R01-NS056114 from the National Institutes of Health.

REFERENCES

1. Williams, A. J., and H. L. Paulson. 2008. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends Neurosci.* 31:521–528.
2. Ross, C. A. 1995. When more is less: pathogenesis of glutamine repeat neurodegenerative diseases. *Neuron.* 15:493–496.
3. Ross, C. A., and M. A. Poirier. 2004. Protein aggregation and neurodegenerative disease. *Nat. Med.* 10:S10–S17.
4. Chen, S. M., V. Berthelier, J. B. Hamilton, B. O'Nuallain, and R. Wetzel. 2002. Amyloid-like features of polyglutamine aggregates and their assembly kinetics. *Biochemistry.* 41:7391–7399.
5. Perutz, M. F., J. T. Finch, J. Berriman, and A. Lesk. 2002. Amyloid fibers are water-filled nanotubes. *Proc. Natl. Acad. Sci. USA.* 99:5591–5595.
6. Sikorski, P., and E. Atkins. 2005. New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils. *Biomacromolecules.* 6:425–432.
7. Chen, S., V. Berthelier, W. Yang, and R. Wetzel. 2001. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.* 311:173–182.
8. Masino, L., G. Kelly, K. Leonard, Y. Trottier, and A. Pastore. 2002. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett.* 513:267–272.
9. Crick, S. L., M. Jayaraman, C. Frieden, R. Wetzel, and R. V. Pappu. 2006. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. USA.* 103:16764–16769.
10. Vitalis, A., X. Wang, and R. V. Pappu. 2007. Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys. J.* 93:1923–1937.
11. Chen, S. M., F. A. Ferrone, and R. Wetzel. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. USA.* 99:11884–11889.
12. Vitalis, A., X. Wang, and R. V. Pappu. 2008. Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *J. Mol. Biol.* 384:279–297.
13. Vitalis, A., and R. V. Pappu. 2009. ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J. Comput. Chem.* 30:673–700.
14. Bhattacharya, A. M., A. K. Thakur, and R. Wetzel. 2005. Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction. *Proc. Natl. Acad. Sci. USA.* 102:15400–15405.
15. Engh, R. A., and R. Huber. 1991. Accurate bond and angle parameters for x-ray protein-structure refinement. *Acta Crystallogr. A.* 47:392–400.
16. Kaminski, G. A., R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B.* 105:6474–6487.
17. Still, W. C., A. Tempczyk, R. C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.
18. Gallicchio, E., and R. M. Levy. 2004. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* 25:479–499.
19. Feig, M., W. Im, and C. L. Brooks. 2004. Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.* 120:903–911.
20. Lazaridis, T., and M. Karplus. 1999. Effective energy function for proteins in solution. *Proteins Struct. Funct. Genet.* 35:133–152.
21. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. 1992. The weighted histogram analysis method for

free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–1021.

22. Roux, B. 1995. Calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* 91:275–282.
23. Hobohm, U., M. Scharf, R. Schneider, and S. Sander. 1992. Selection of a representative set of structures from the Brookhaven Protein Data Bank. *Protein Sci.* 1:409–417.
24. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.
25. Sugita, Y., A. Kitao, and Y. Okamoto. 2000. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* 113:6042–6051.
26. Camacho, C. J., and D. Thirumalai. 1993. Minimum energy compact structures of random sequences of heteropolymers. *Phys. Rev. Lett.* 71:2505–2508.
27. Milchev, A., W. Paul, and K. Binder. 1993. Off-lattice Monte Carlo simulation of dilute and concentrated polymer solutions under theta conditions. *J. Chem. Phys.* 99:4786–4798.
28. Marchut, A. J., and C. K. Hall. 2007. Effects of chain length on the aggregation of model polyglutamine peptides: molecular dynamics simulations. *Proteins Struct. Funct. Bioinform.* 66:96–109.
29. Barton, S., R. Jacak, S. D. Khare, F. Ding, and N. V. Dokholyan. 2007. The length dependence of the polyQ-mediated protein aggregation. *J. Biol. Chem.* 282:25487–25492.
30. Lee, C. C., R. H. Walters, and R. M. Murphy. 2007. Reconsidering the mechanism of polyglutamine peptide aggregation. *Biochemistry.* 46:12810–12820.
31. Krishnan, R., and S. L. Lindquist. 2005. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature.* 435:765–772.
32. Lomakin, A., N. Asherie, and G. B. Benedek. 2003. Liquid-solid transition in nuclei of protein crystals. *Proc. Natl. Acad. Sci. USA.* 100:10254–10257.
33. Pappu, R. V., X. Wang, A. Vitalis, and S. L. Crick. 2008. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch. Biochem. Biophys.* 469:132–141.
34. Bitan, G., M. D. Kirkitadze, A. Lomakin, S. S. Vollers, G. B. Benedek, et al. 2003. Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. USA.* 100:330–335.
35. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD—visual molecular dynamics. *J. Mol. Graph.* 14:33–38.