JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

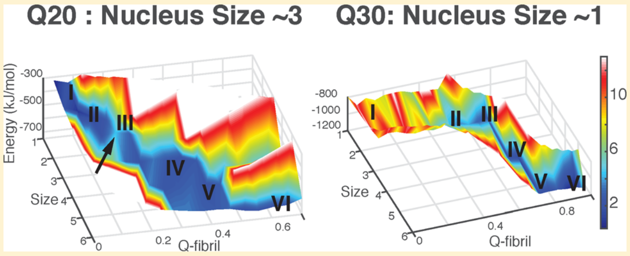
The Aggregation Free Energy Landscapes of Polyglutamine Repeats

Mingchen Chen,[†,‡] MinYeh Tsai,[†,§] Weihua Zheng,[†,§] and Peter G. Wolynes*[,†,§]

[†]Center for Theoretical Biological Physics, [‡]Department of Bioengineering, and [§]Department of Chemistry, Rice University, Houston, Texas 77005, United States

Supporting Information

ABSTRACT: Aggregates of proteins containing polyglutamine (polyQ) repeats are strongly associated with several neurodegenerative diseases. The length of the repeats correlates with the severity of the disease. Previous studies have shown that pure polyQ peptides aggregate by nucleated growth polymerization and that the size of the critical nucleus (n*) decreases from tetrameric to dimeric and monomeric as length increases from Q_{18} to Q_{26}. Why the critical nucleus size changes with repeat-length has been unclear. Using the associative memory, water-mediated, structure and energy model, we construct the aggregation free energy landscapes for polyQ peptides of different repeat-lengths. These studies show that the monomer of the shorter repeat-length (Q_{20}) prefers an extended conformation and that its aggregation indeed has a trimeric nucleus (n* ∼ 3), while a longer repeat-length monomer (Q_{30}) prefers a β-hairpin conformation which then aggregates in a downhill fashion at 0.1 mM. For an intermediate length peptide (Q_{26}), there is an equal preference for hairpin and extended forms in the monomer which leads to a mixed inhomogeneous nucleation mechanism for fibrils. The predicted changes of monomeric structure and nucleation mechanism are confirmed by studying the aggregation free energy profile for a polyglutamine repeat with site-specific PG mutations that favor the hairpin form, giving results in harmony with experiments on this system.



1. INTRODUCTION

The origin of at least eight neurodegenerative diseases can be genetically traced to the presence of proteins having long repeats of polyglutamine (polyQ).[1] The severity of these diseases typically depends on the polyQ repeat-length. One of these afflictions, Huntington's disease (HD) arises only in patients having proteins with 36 repeats or more; and individuals with genes coding for longer repeats acquire symptoms earlier.[2] It has been established that the rate of aggregation increases with the repeat-length both in vivo and in vitro.[3]

While glutamine is amphiphilic, the aggregates of polyQ appear to be amyloids having the β-strand as their basic unit.[4−8] Why such amphiphilic sequences that do not conform to the usual hydrophobic patterns of most amyloidogenic fragments should aggregate has been mysterious.[9] Primary nucleation has been considered to be the rate-limiting step in polyQ aggregation.[10,11] Experiments on pure polyQ repeat peptides show that the critical nucleus size decreases from n* = 4 to 1 as the repeat-length grows from Q_{18} to Q_{26}.[3] Simulations of protein aggregation remain challenging even with coarse-grained models,[12] as most of the coarse-grained models suffer from the question of realism to fold globular proteins even when they are adequate to capture many interesting polymer physics aspects of aggregation. To address the origin of this biophysically striking length dependence of aggregation, in this paper, we explore the free energy landscapes for aggregation of oligomers of polyQ sequences having different repeat-lengths using the coarse-grained associative memory, water-mediated, structure and energy model (AWSEM) force field. The AWSEM force field, while being efficient to simulate, also has been shown to predict the structures of protein monomers[13,14] as well as details of their assembly.[15−17] We have previously used the force field to study the free energy landscape of a mechanical prion, also rich in glutamine, that is involved in memory.[17] We find that while the simple polyQ peptides longer than 30 residues prefer a hairpin structure in the monomer, for shorter lengths an extended structure is favored. The critical nucleus size inferred from the simulated aggregation free energy landscapes for the different length peptides agrees well with the laboratory results. At the laboratory concentration of peptide, the free energy profile for Q_{20} shows a nucleation barrier near the trimer, while the Q_{30} peptide aggregates in a downhill fashion at the same concentration. For a polyQ peptide of intermediate length (Q_{26}), inhomogeneous nuclei for aggregation are observed. These nuclei lead to branched structures rather than simple one-dimensional (1D) fibers. The simulations suggest that the mixed, inhomogeneous character of the nucleation species for the intermediate length peptide is the cause of this "branching" behavior. These results suggest a way in which fiber growth may be controlled with point mutations in peptide sequence.

Received: August 18, 2016
Published: October 27, 2016

DOI: 10.1021/jacs.6b08665
J. Am. Chem. Soc. 2016, 138, 15197−15203

2. METHODS

2.1. The Abbreviated AWSEM Force Field. A detailed description of the AWSEM force field has already been given in Davtyan et al.[13] Briefly, AWSEM is a predictive coarse-grained protein folding force field that employs three sites per amino acid whose parameters are learned from structural data using an algorithm based on energy landscape theory.[18] The AWSEM Hamiltonian summarized in eq 3 consists of a backbone term V_{backbone}, a many body burial term V_{burial}, a contact term V_{contact}, and a hydrogen-bonding term V_{HB}.

The amino acid residues in the AWSEM force field are all encoded to adopt the L-configuration through the chirality term, V_χ, inside the backbone term. The chirality term, given in the following equation, ensures the orientation of the C_β atom relative to the plane formed by the C', C_α, and N atoms. The value $\chi_0 = -0.83$ Å^3 corresponds to the L-type amino acid.[13]

$$V_\chi = \lambda_\chi \sum_{i=2}^{N-1} (\chi_i - \chi_0)^2 \tag{1}$$

$$\chi_i = (r_{C_i'C_{\alpha i}} \times r_{C_{\alpha i}N_i})\cdot r_{C_{\alpha i}C_{\beta i}} \tag{2}$$

The hydrogen-bonding term, V_{HB}, is associated with the secondary structural weight in AWSEM which encodes a bias for the α-helix conformation and another bias for β-strand hydrogen-bond formation. While the strength for these terms for normal globular proteins is based on secondary structural prediction, for intrinsically disordered proteins like polyQ, secondary structure prediction may not be adequate. We have tuned the strength in our previous paper on the CPEB protein to study the aggregation of Q-rich proteins.[17] In that analysis, we determined the stability of secondary structures with different secondary structural bias weights and compared the stability from the AWSEM simulations with the statistics from separate all-atom simulations. A set of secondary structural bias weights was then chosen so that the relative stability of different structures is similar for both the all-atom simulations and the AWSEM simulations. Details may be found in the SI of ref 17. We use the same weights here.[17]

Additional local-in-sequence interactions governed by the fragment associative memory term, V_{FM}, are ordinarily determined by bioinformatic sequence matching.[13] These interactions are normally useful for structural prediction. These are not included in the force field used in the present study, because very few local sequences of polyQ have been structurally characterized in the Protein Data Bank.[14] Such sequences are usually classified as being "intrinsically disordered". We therefore call the present version of the force field, abbreviated AWSEM.

$$V_{\text{total}} = V_{\text{backbone}} + V_{\text{contact}} + V_{\text{burial}} + V_{\text{HB}} \tag{3}$$

The full AWSEM force field has already been used with much success in predicting globular protein structures for both monomers and dimers. It has also been used to study the initial stages of misfolding and aggregation.[15,16] We have previously used the abbreviated AWSEM force field to explore the aggregation of another Q-rich peptide CPEB, which has been postulated to be a mechanical prion involved in memory.[17]

2.2. Order Parameters for Umbrella Sampling and Free Energy Calculations. To survey the energy landscape in a quantitative way, it is useful to employ a variety of collective order parameters to classify structural ensembles. The structural similarity of two different protein configurations is generally characterized by a local order parameter, $Q^{\alpha\beta}$, which is given by the equation:

$$Q^{\alpha\beta} = \frac{2}{(N-2)(N-3)} \sum_{j>i+2} \exp[-(r_{ij}^{\alpha} - r_{ij}^{\beta})^2/2\delta_{ij}^2)] \tag{4}$$

where N is the total number of residues. When a folded structure is known (or postulated), it is useful to choose one of the structures in the comparison to be the (assumed) native structure. We then call $Q^{\alpha\beta}$ simply Q.

We have used umbrella sampling to enhance the sampling of conformations for free energy calculations. A harmonic potential in Q with respect to a final fibril structure found in our simulated annealing runs is used to restrain constant temperature molecular dynamics simulations to give configurations within a range of reference values:

$$V_{Q-\text{bias}} = \frac{1}{2}k_{Q-\text{bias}}(Q - Q_0)^2 \tag{5}$$

with $k_{Q-\text{bias}} = 200$ kcal/mol. The reference values for Q_0 are chosen to be equally spaced from 0 to 0.98 with a step size 0.02. Data from different windows are stitched together using the weighted histogram analysis method (WHAM) to construct full free energy landscapes.

2.3. Simulation Details. All simulations were performed using the software of the large-scale atomic/molecular massively parallel simulator (LAMMPS). The AWSEM force field is now available in this open source format.[13] All the umbrella sampling simulations for multiple peptide chains are performed in a cubic box (500 A × 500 A × 500 A) with periodic boundary condition for 20 million steps at 370 K. The initial configurations of the umbrella sampling for oligomer simulations are six monomers randomly distributed over the cubic box to give a nominal concentration of 0.1 mM. The peptides are started in extended conformations. In fact monomer configurations equilibrate rapidly. Because of this, the details of these initial conformations are erased by thermal motions long before aggregation begins. Free energy profiles are calculated and extrapolated to the physiological temperature (300 K) using the weighted histogram analysis method (WHAM). We run the umbrella sampling at a higher temperature where oligomers readily form and disassociate. Simulation at 370 K enhances considerably the sampling efficiency. The extrapolation to physiological temperature is indeed rather short and should not affect the results very much.

2.4. Correction for the Concentration of Free Monomers. During simulations, as aggregates form, the concentration of free monomers changes because of the small size of the simulation box. This change must be accounted for in order to relate simulations to macroscopic experiments that are carried out at essentially fixed laboratory concentrations and chemical potentials. We describe in this section how to carry out the corrections for this effect. Additional details on the strategy may be found in our earlier paper on amyloid-β aggregation.[19]

Suppose that a system has a total of N monomers in a volume V at temperature T and that some of these n monomers aggregate to form oligomers. If only one oligomer of size n is formed, then there are the $N - n$ free monomers remaining in solution once the single oligomer has formed. The probability that there is at least one n-oligomer in the system according to Reiss and Bowles's approach[20] is given by

$$\begin{aligned} P_n &= \frac{Q_n(N, V, T)}{Q(N, V, T)} = \frac{Q(N-n, V, T)q_n(V, T)}{Q(N, V, T)} \\ &= q_n(V, T)e^{n\mu/k_BT} \end{aligned} \tag{6}$$

where $Q_n(N, V, T)$ denotes the partition function of the system constrained to contain at least one n-oligomer and that $Q(N, V, T)$ represents the total unconstrained partition function; $q_n(V, T)$ is the partition function of the oligomer in which the interaction with the remaining $N - n$ monomers is included; $Q(N - n, V, T)$ is the partition function for the decoupled remaining $N - n$ monomers; and μ is the chemical potential of a monomer. In the derivation leading to eq 6, the following equation is used:

$$Q_n(N, V, T) = Q(N-n, V, T)q_n(V, T) \tag{7}$$

which provides an analytical scheme so as to decouple the configuration integral for the n-oligomer ($q_n(V, T)$) from the remaining $N - n$ monomers, denoted by $Q(N - n, V, T)$. The potential energy includes three components: (1) the interactions within the n-oligomer; (2) the interactions between the n-oligomer and the remaining $N - n$ monomers; and (3) the interactions within the remaining $N - n$ monomers. Note that the internal structural details are irrelevant to the theoretical development of the correction factors. This theoretical approach is basically the same as that employed in the physical cluster approach to the nucleation of liquids from gases.[21]

Now, given the fact that in our simulations a large n-oligomer is rare in the small simulation box, P_n is roughly same as the probability that there is exactly one n-oligomer in the system. It follows that

$$F_n = -k_B T \ln P_n = -k_B T \ln q_n - n\mu \tag{8}$$

This result is used to compute the thermodynamic potential for the grand canonical ensemble $F_n - n\mu$. The gradients of this potential determine the growth or dissociation of clusters of size n in a system of fixed chemical potential μ.

When we get the free energy for the grand canonical ensemble $F_n - n\mu$ at concentration C_0, we can extrapolate the free energy to a different concentration C_1 by using eq 9:

$$\begin{aligned} F_n(C_1) &= F_n - n\mu_1 = F_n(C_0) - n(\mu_1 - \mu_0) \\ &= F_n(C_0) - nkT \ln \frac{C_1}{C_0} \end{aligned} \tag{9}$$

where μ_1 and μ_0 are the chemical potentials of the free monomers at concentration C_1 and C_0, respectively.

3. RESULTS

3.1. The Structural Preference of Monomeric PolyQ for Being an Extended β-Strand versus Being a β-Hairpin Depends on Length. In solution monomeric polyQ peptides have been investigated both experimentally and computationally by many groups leading to their being classified as "intrinsically disordered".[22−25] Circular dichroism (CD) suggests that polyQ peptides are largely disordered and gives little hint of there being any difference between the short polyQ sequences and longer ones.[23,26−28] The CD analysis suggests that Q_{40} has about 10% α-helix content and only a trace amount of β-sheet and β-turn.[29] The solved crystal structure of the Huntingtin exon1 Htt17Q-EX1 fused with maltose binding protein, the exon itself consisting of an amino-terminal α-helix, poly17Q, and a polyproline 3−10 helix, suggests that the polyQ region itself adopts multiple conformations, including α-helix, random coil, and extended loop.[24]

Finding the structure of the aggregated form long has been a challenge. Perutz first proposed the repeats could form a β-helix, whose helical repeat would offer an explanation for how length could affect the age-of-onset.[30] More recent X-ray analyses, ss-NMR data, and mutational studies, however, argue against this structural picture and instead favor a standard β-strand model for the aggregate.[4−8] The stabilities of different structures of polyQ have also been addressed computationally using all-atom simulations. These suggest the β-hairpin model is the most stable form for Q_{40}.[31]

We study primarily in this paper the pure polyglutamine repeat peptides, Q_{20}, Q_{24}, Q_{26}, Q_{30}, and Q_{40}, using the abbreviated AWSEM force field. For each monomer, we compute 1D free energy profiles using umbrella sampling with respect to an order parameter that measures generic hairpin formation: the number of contacts formed by each residue. The hairpin structure (having an average number of contacts per unit length around 0.45) dominates thermodynamically over the extended structure (with an average number of contacts per unit length around 0) for monomeric Q_{30} and Q_{40}. But in Q_{20}, the extended structure is favored(Figure 1). For Q_{24} and Q_{26}, the two classes of conformations are equally favorable. For the longest peptide that we studied, polyQ (Q_{50}), the most favorable state turns out to be a three-strand β hairpin structure (see Figure S1). Hairpin states form much more easily in longer repeats. Consistent with the proposed idea that the hairpin is the building block for the fiber form, we see the trend of monomer structure alone suggests an explanation for the length dependence of polyQ aggregation. We wish to remind the reader that for all lengths of peptide, the thermal ensemble is rather broad, as evidenced by the modest free energy barriers seen in Figure 1. Nevertheless the aggregation mechanism can be understood using the structural preferences described by these two configurational classes.

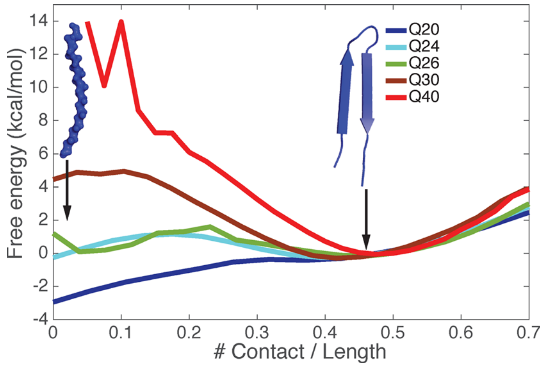


Figure 1. Free energy profiles for monomers of different lengths reveal their different structural preferences between extended β-strand and β-hairpin. Results for polyQ peptides of different lengths are color-coded.

3.2. The Calculated Critical Nucleus Size (n*) for Q_{20} is 3−4. The rate of aggregation depends on concentration. The grand canonical potential for following the aggregation process is $F_n - n\mu$, where μ is the chemical potential set by the monomer concentration. At very low concentration, $F_n - n\mu$ monotonically increases with n favoring the monomer, but when the solution is supersaturated, $F_n - n\mu$ eventually decreases with n favoring a large aggregate. In between, there may be a nucleation barrier at a finite number of monomers aggregated into a unit. If n, the aggregate size, is a sufficiently good reaction coordinate, then the peak in $F_n - n\mu$ indicates the size of the critical nucleus, which in turn determines the concentration dependence of the aggregation rate near that chemical potential. If the critical nucleus size is n*, the nucleation rate will vary as C^{n*}, where C is the concentration.

Experiments show that the critical nucleus size (n*) of polyQ is length dependent. To compute the free energy profile at a fixed laboratory concentration and a physiological temperature, we first compute the aggregation free energy for the entire simulation box as a function of the total number of inter-residue contacts in a system containing six Q_{20} monomers (nominal concentration ~0.1 mM) at 300 K. As shown in the 1D free energy plot at this nominal concentration, the free energy of formation of an oligomer is uphill until four units are assembled, while it is downhill afterward. We visualized structures in each of the contact number basins. As shown in Figure 2A, each basin corresponds with a different size oligomer. This free energy profile versus contact number suggests that the critical nucleus size is around 4 at the nominal simulation concentration (Figure 2A). The formation of larger oligomers of Q_{20} is not favored in the simulation with fixed box size and total monomer count. The aggregation process is energetically favored (Figure 2B), since the potential energy decreases monotonically as the oligomer grows, but the entropy

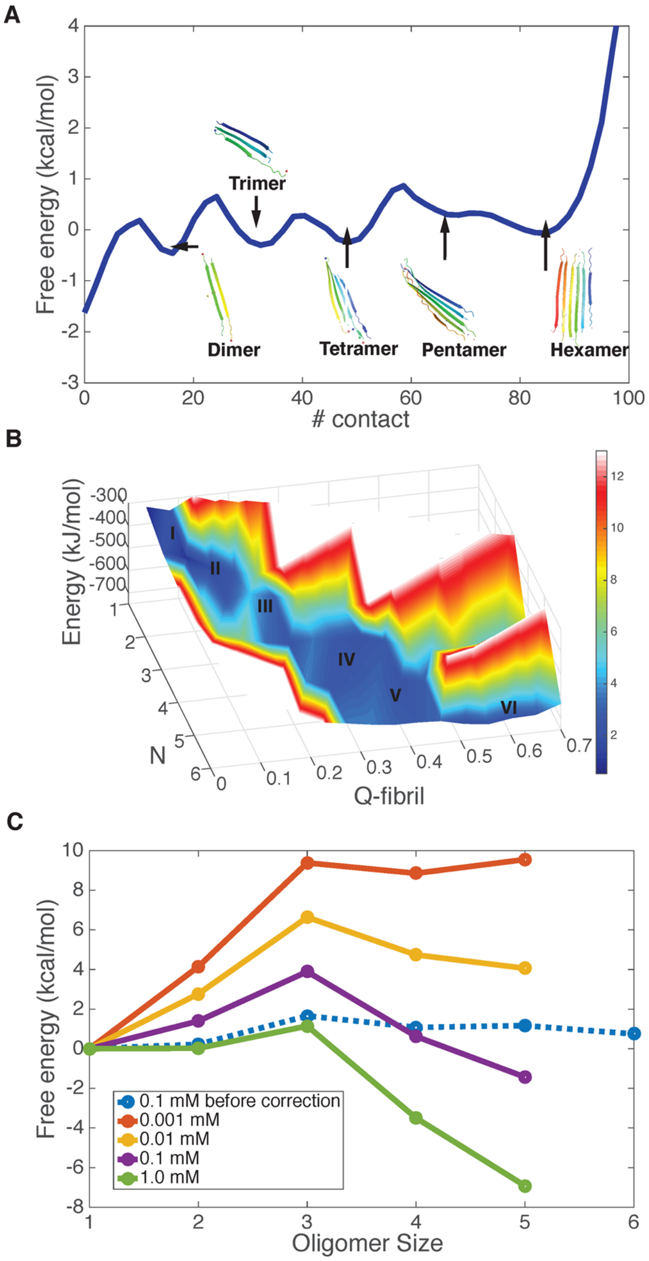


Figure 2. Aggregation free energy landscape for Q_{20} at 300 K. (A) The free energy profile as a function of the number of total residue−residue contacts in the simulation system with six Q_{20} peptide chains. Representative structures of different oligomers are shown in each free energy basin of the progress in different oligomeric states. (B) The energy and free energy surfaces for aggregation of Q_{20} are plotted as a function of the oligomer size (N), along with its structure similarity compared to the final fiber form (Q_{fibril}). The z-axis is the energy of the system, which decreases monotonically as the oligomer size increases. The color indicates the free energy. The local basins for different oligomer states are labeled by size. A free energy barrier occurs around $N = 3$. (C) The grand canonical free energy for different oligomer states as corrected for concentration changes shows the saturation value of the concentration of free monomers.

cost of adding monomers to the large cluster is too large once the monomers in the simulation box are depleted.

To exactly compute the critical nucleus size at a fixed laboratory concentration and the physiological temperature, we must calculate the free energy as a function of the size of the oligomer (color red in Figure 2C) but keep the concentration for free monomers fixed. In the simulations with a finite box, however, the monomers are depleted as the large cluster grows, an effect that is negligible in the laboratory. Fortunately a correction for this effect can be carried out using the Reiss physical cluster approach as described in the Methods section. After the correction for the change in concentration of free monomers, the grand canonical free energy profile, $F_n - n\mu$, shows a peak at size 3 for the laboratory concentration of 0.1 mM, suggesting that the critical nucleus size measured under these conditions should be 3 (Figure 2C). Experiments for Q_{20} at the same concentration give a somewhat larger nucleus n ∼ 4. When we extrapolate the concentration to 1 μM, the aggregation free energy profile is uphill; while the profile becomes downhill asymptotically at a strong supersaturation (1 mM). The critical saturation concentration (c*) is therefore predicted to be between 10−100 μM. This prediction compares well with experiments that show that a polyglutamine repeat of a similar length of 23 does not aggregate at 52 μM, but that aggregation proceeds slowly at the concentration of 103 μM[3] at the physiological temperature.

The structures of the pentamers and hexamers found in our simulation resemble the models proposed by Schneider et al.[5] These structures suggest the fiber form will be composed simply of layers of extended β-strands when the repeat-length is short.

3.3. The Calculated Critical Nucleus Size (n*) for Q_{30} Is 1.

The simulations suggest the extended structure is the preferred form of the monomer for Q_{20}, and the critical nucleus size for aggregation is ∼3−4. We now examine the aggregation for Q_{30} whose monomers show a preference for the hairpin structure. Again we perform simulations with six peptide chains in a box (the nominal concentration is ∼0.1 mM) and compute the free energy profile for forming aggregates of Q_{30} at the physiological temperature. Now the free energy profile as directly simulated (even with monomer depletion) is almost downhill when we use the total number of contacts in the simulation cell as the reaction coordinate. The profile exhibits only a single peak which occurs when the hairpin in the monomers is formed from an extended state. The small free energy (∼$2k_B$T) barrier that occurs during the restructuring of the monomeric hairpin suggests the critical nucleus size as measured by concentration dependence would appear to be 1 for Q_{30} (Figure 3A). The free energy barrier during hairpin formation is indeed the rate-limiting step in aggregation of longer polyglutamine repeats. As we increase the repeat length, the barrier starts to vanish (Figure 1), which leads to the higher aggregation rates observed in experiments.[32]

The free energy profile after making the correction for the changing concentration of free monomers in the simulation exhibits an even more downhill behavior (Figure 3C), in harmony with the relatively high aggregation rate of Q_{30} in vitro. This sort of behavior is referred to by Ferrone as a "monomeric" aggregation nucleus.[33] When we extrapolate the concentration to 1 μM, there is a barrier in the aggregation process; while it becomes much more downhill under supersaturation (1 mM). The predicted critical concentration (c*) is between 1 and 10 μM, and this value compares well with experiments showing that solubility limit for $Q_{30}K_2$ is around 5 μM at 303 K.[34] The most favorable oligomeric structures from the free energy basin are antiparallel β-hairpin sheets. This structural prediction agrees well with what has been inferred from experimental ssNMR on aggregates and point mutation studies.[8] We did not observe in our simulations either the

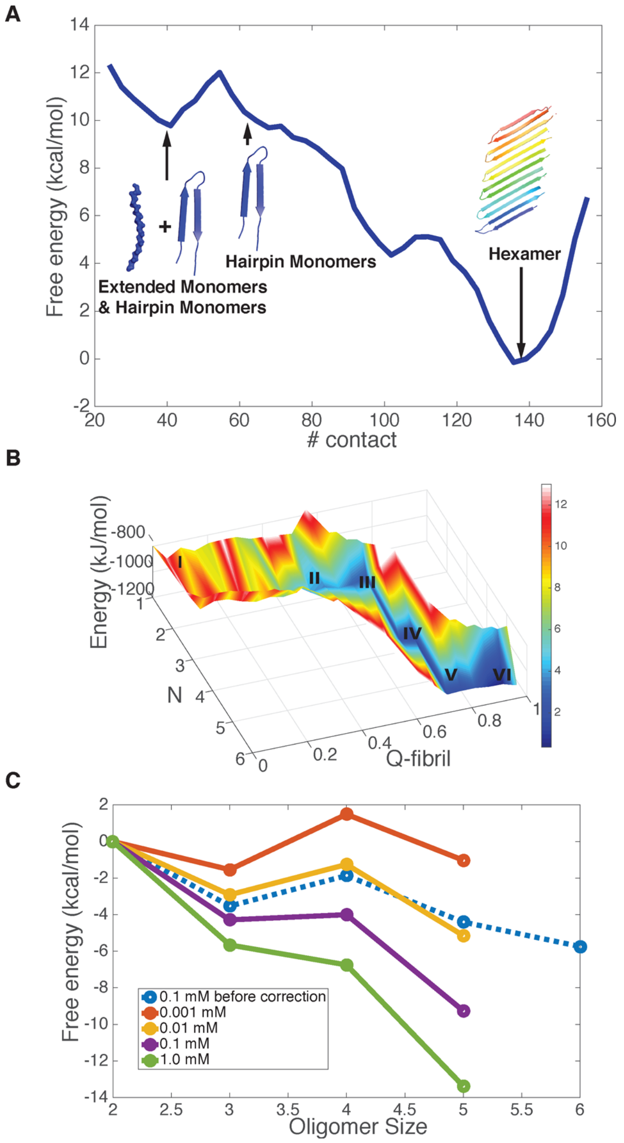


Figure 3. Aggregation free energy landscape for Q_{30} at 300 K. (A) The free energy profile as a function of the number of total residue−residue contacts in the finite size simulation system with six Q_{30} peptide chains. Representative structures in each basin illustrate the progression through the different oligomeric states. (B) The energy and free energy surfaces for aggregation of Q_{30} are plotted as a function of the oligomer size (N), and its structure similarity compared to the final fiber form (Q_{fibril}). The z-axis is the energy of the system, and it decreases monotonically as the oligomer size increases. The color indicates the free energy, which includes the entropy cost of addition at concentration 0.1 mM. The local basins for different oligomer states are labeled by size. (C) The grand canonical free energy for different oligomer states as corrected for the finite size effect is shown for different concentrations of free monomers.

parallel β-hairpin structures or the β-arch structures which had been postulated by others.[5,35] We also notice that other forms of β-structures like β-arch have been observed computationally,[36] employing much more crudely coarse-grained simulations, in comparison with the AWSEM model. In contrast, the AWSEM prediction of the fibril structure compares very well with the 2D-FTIR results from Zanni's group.[7]

3.4. "Mixed or Inhomogeneous" Nucleation for Q_{26} Leads to Branched Filaments.

As the polyQ length increases from 20 to 30 units, the hairpin structure begins to dominate over the extended form in the monomer, and the critical nucleus size for aggregation decreases from 4 to 1. For an intermediate length peptide, Q_{26}, the simulation results suggest that both the extended form and hairpin form of the monomer are equally favorable. In this case, how does aggregation occur? We again compute the 1D free energy profile at the physiological temperature (Figure S2A) using the total number of contacts in the simulation box as the reaction coordinate. This profile shows a peak in the formation of hairpin structures from extended form, and the critical nucleus size is 1 at 300 K. But the most favorable final fiber form structure turns out to be a mixture of hairpins and extended chains (Figure S2A), instead of the simple pure antiparallel hairpins formed for Q_{30}. As shown in the 1D free energy plot as a function of the size of oligomer, the free energy of forming an oligomer is downhill before forming a desired antiparallel β-hairpin hexamer, while forming a desired antiparallel β-hairpin hexamer is uphill (Figure S2B). The free energy profile after making the correction for changing the concentration of free monomers in the simulation exhibits a downhill behavior (Figure S2B). The predicted critical concentration (c^*) is between 1 and 10 μM at the physiological temperature (refer to the SI for details).

We observed several different intermediate structures during nucleation simulations (Figure 4A). These turned out to be a mixture of β-hairpins and extended β-strands. The calculations suggest the mechanism for aggregation will have a mixed or inhomogeneous set of nuclei (Figure 4B). The variety of distinct nuclei comes from there being two equally favorable conformations for Q_{26}. As the chain grows, the extended monomers can attach onto the hairpins and vice versa. This attachment widens the fiber in a direction perpendicular to the main fiber axis. We call this mechanism "branching". Walters et al. have shown in their experiments that while polyQ sequences of a shorter length and those of a longer length form fibers with clear bundles, the fiber morphology of the intermediate length peptides Q_{26} shows a lateral alignment.[25] We see that this lateral alignment phenomenon is explained by the present simulations.

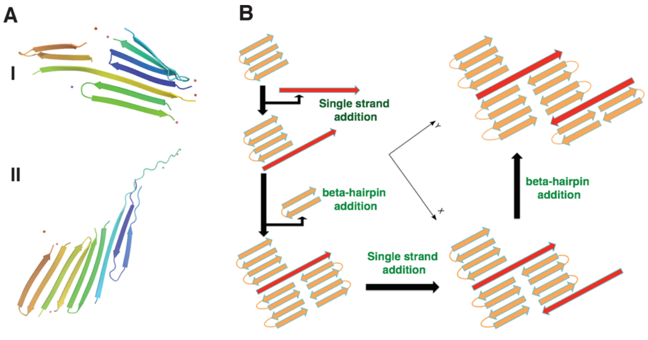


Figure 4. Proposed mechanism for the formation of mixed nuclei in the aggregation of Q_{26}. (A) Selected representative structures for Q_{26} with different topologies. (B) A schematic diagram showing the of formation of "branched" oligomers from mixed nuclei. The different monomer components are color-coded, with brown representing the hairpin structure and red indicating the peptide takes on an extended structure.

3.5. $Q_{12}PGQ_{12}$ Nucleation.

The possible involvement of the hairpin in the nucleation mechanism had already suggested studying the aggregation of peptides that encourage the formation of hairpins.[8,25,37]. Wetzel and co-workers have constructed peptides in which Q's in the middle of the sequence are mutated to proline and glycine (PG mutation). Such a mutation stabilizes a hairpin at the location of the PG kink. They have shown that Q_9PGQ_9 now already shows downhill aggregation behavior.[8,25] Further more, the PG mutation eliminates the lateral alignment behavior in fiber morphology.[25]

We examined these same peptides using our simulation approach. We computed the free energy profiles for $Q_{12}PGQ_{12}$ monomers. We see that there is a very much enhanced β-hairpin propensity (Figure 5A). Now already at length 26, the free energy difference favors the hairpin by about 6 kcal/mol. The aggregation free energy profile at the physiological temperature without the depletion correction for $Q_{12}PGQ_{12}$ shows the nucleation process is downhill, with the antiparallel hairpin fiber becoming the most stable final state (Figure 5B). After making a depletion correction, the free energy profile is downhill (Figure 5C). The predicted critical concentration is between 1 and 10 μM at 300 K, agreeing perfectly with what is found for the polyglutamine peptide PG mutations using L-amino acids($Q_{10}PGQ_{11}$).[8] The L-PG mutation does not enhance the aggregation rate significantly, but the mutation facilitates β-hairpin formation in the aggregates leading to a "monomeric nucleus" with n* ~ 1.[8] Kar et al. have also studied the PG mutations to D-form amino acids. This mutation not only leads to n* ~ 1 but also enhances the aggregation with a critical concentration of 0.4 μM.[8] We did not carry out any simulations of this system with its backbone of unnatural chirality, which requires modification of the abbreviated AWSEM software.

4. DISCUSSION

4.1. The Persistence Length of PolyQ Peptides Explains Their Structural Preferences. Our simulations suggest that the monomer's structural preferences do greatly influence fiber nucleation and growth. Polymer physics provides an explanation for the structure change with length. Increasing the chain length increases the probability for collapse. The persistence length of polyQ in solution corresponds to ~3–4 residues in sequence.[23] Two parts of a peptide chain can only make contacts when they are separated by at least 4 persistence lengths.[38] In Q_{20} (4–5 persistent lengths), the probability to make contacts and collapse is still small, but the collapse probability is much larger for Q_{40}.

4.2. Why Does the Nucleus Possess a Similar Structure to the Proposed Fiber Form? In our recent work on the aggregation of Aβ40, we found the oligomers of small size must reconfigure by rearranging before they can form the most stable fiber (i.e., they "backtrack"). The Aβ40 first assembles into prefibrilar forms which are thought by some to be the key to the pathology, and then these prefibrilar oligomers rearrange into parallel β-hairpin structures. In contrast, for polyglutamine repeats, the nucleus already resembles the final β-strand structure of the fiber form. Unlike what happened for Aβ40, the monotonous nature of the sequence leads to a relatively simple aggregation mechanism that underlies the simple length dependence of fiber nucleation and assembly.

The situation is more complicated for the in vivo aggregation of the species involved in HD, peptides based on Huntingtin exon1. Recent studies on the full-length Huntingtin exon1 do find evidence of oligomers of heterogeneous structures. These small oligomers have been proposed to be the pathogenic culprits in this case too.[39−41] We are currently studying the aggregation of Huntingtin exon1 peptide and plan to report on our results for these much larger systems later.

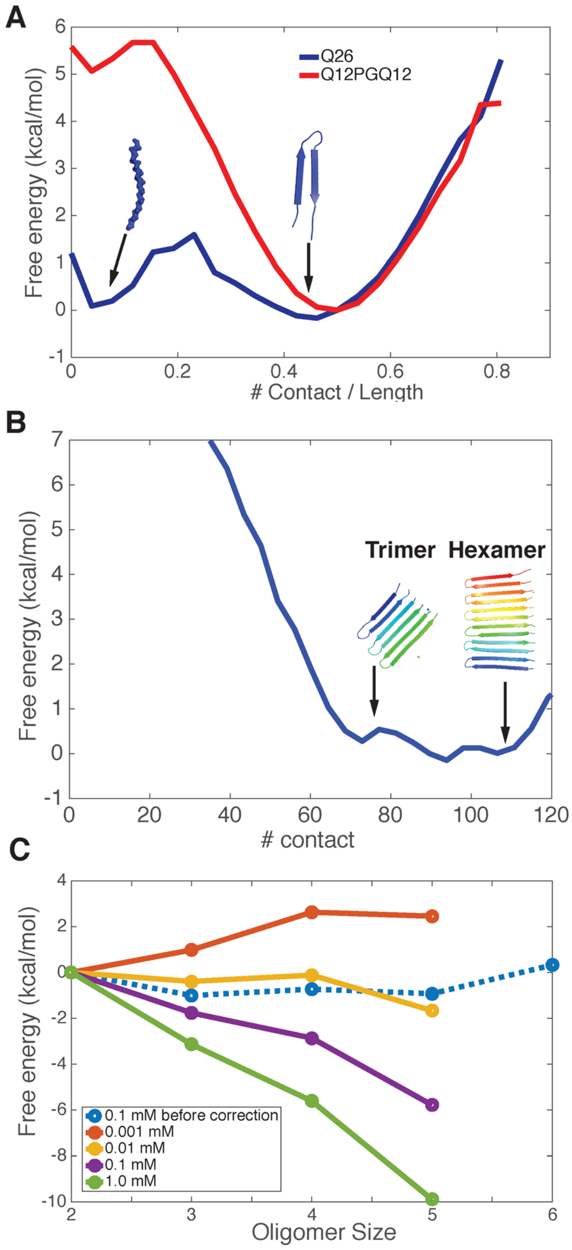


Figure 5. Changing the structural preference for 26 length repeats by making point mutations alters the nucleation behavior at 300 K. (A) An inserted L-PG mutation makes the hairpin conformation more favorable. (B) The free energy profile as a function of the number of total residue−residue contacts in a simulation system with six $Q_{12}PGQ_{12}$ peptide chains. Representative structures in each basin illustrate the progression through the different oligomeric states. (C) The grand canonical free energy for different oligomer states as corrected for the finite size effect for different concentration of free monomers.

4.3. The Branching Nucleation Behavior Suggests a Way To Control Fiber Morphology. In our simulations, when the polyglutamine repeat length is either rather short or rather long, additional β-strands would attach to the existing aggregate primarily by following the direction of the fiber axis. At intermediate length, attaching the two equally favorable conformations of the monomer to the aggregate gives rise to a "branching" mechanism which leads to lateral alignment. The aggregation free energy landscape suggests that the introduction of the PG mutations would eliminate the lateral alignment, and indeed this is seen. These results together give us confidence that simulations can guide the control of fiber morphology in vitro. Fiber growth control has long been a problem in macromolecular engineering and tissue engineering. The morphology of DNA fibers has been carefully controlled using "DNA bricks", and the size of such fibers can reach beyond the nanometer level.[42] By mutating polyglutamine sequences, it should be possible to tune the preference for different monomer conformations, thus influencing fiber morphology. A protein gel made up of different polymeric β-strands may allow the modification of the gel's mechanical properties, which may be of practical use.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b08665.

Experimental details and data (PDF)

AUTHOR INFORMATION

Corresponding Author

*pwolynes@rice.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grant R01 GM44557 from the National Institute of General Medical Sciences. Additional support was also provided by the D. R. Bullard-Welch Chair at Rice University, gGrant C-0016. We thank the Data Analysis and Visualization Cyberinfrastructure funded by National Science Foundation Grant OCI-0959097. M.Y.T. is also supported by Ministry of Science and Technology (MOST), Taiwan, R.O.C., grant number: 103-2917-I-564-015.

REFERENCES

(1) Zoghbi, H. Y.; Orr, H. T. *Annu. Rev. Neurosci.* **2000**, *23*, 217−247.

(2) Walker, F. O. *Lancet* **2007**, *369*, 218−228.

(3) Kar, K.; Jayaraman, M.; Sahoo, B.; Kodali, R.; Wetzel, R. *Nat. Struct. Mol. Biol.* **2011**, *18*, 328−336.

(4) Sambashivan, S.; Liu, Y.; Sawaya, M. R.; Gingery, M.; Eisenberg, D. *Nature* **2005**, *437*, 266−269.

(5) Schneider, R.; Schumacher, M. C.; Mueller, H.; Nand, D.; Klaukien, V.; Heise, H.; Riedel, D.; Wolf, G.; Behrmann, E.; Raunser, S.; Seidel, R.; Engelhard, M.; Baldus, M. *J. Mol. Biol.* **2011**, *412*, 121−136.

(6) Hoop, C. L.; Lin, H.-K.; Kar, K.; Magyarfalvi, G.; Lamley, J. M.; Boatz, J. C.; Mandal, A.; Lewandowski, J. R.; Wetzel, R.; van der Wel, P. C. A. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 1546−1551.

(7) Buchanan, L. E.; Carr, J. K.; Fluitt, A. M.; Hoganson, A. J.; Moran, S. D.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5796−5801.

(8) Kar, K.; Hoop, C. L.; Drombosky, K. W.; Baker, M. A.; Kodali, R.; Arduini, I.; van der Wel, P. C. A.; Horne, W. S.; Wetzel, R. *J. Mol. Biol.* **2013**, *425*, 1183−1197.

(9) Thirumalai, D.; Reddy, G.; Straub, J. E. *Acc. Chem. Res.* **2012**, *45*, 83−92.

(10) Chen, S.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11884−11889.

(11) Bhattacharyya, A. M.; Thakur, A. K.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15400−15405.

(12) Morriss-Andrews, A.; Shea, J.-E. *Annu. Rev. Phys. Chem.* **2015**, *66*, 643−666.

(13) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. *J. Phys. Chem. B* **2012**, *116*, 8494−8503.

(14) Chen, M.; Lin, X.; Zheng, W.; Onuchic, J. N.; Wolynes, P. G. *J. Phys. Chem. B* **2016**, *120*, 8557−8565.

(15) Zheng, W.; Schafer, N. P.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1680−1685.

(16) Zheng, W.; Schafer, N. P.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 20515−20520.

(17) Chen, M.; Zheng, W.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 5006−5011.

(18) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.

(19) Zheng, W.; Tsai, M.-Y.; Chen, M.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 11835−11840.

(20) Reiss, H.; Bowles, R. K. *J. Chem. Phys.* **1999**, *111*, 7501.

(21) Gillis, H. P.; Marvin, D. C.; Reiss, H. *J. Chem. Phys.* **1977**, *66*, 223.

(22) Wang, X.; Vitalis, A.; Wyczalkowski, M. A.; Pappu, R. V. *Proteins: Struct., Funct., Genet.* **2006**, *63*, 297−311.

(23) Walters, R. H.; Murphy, R. M. *J. Mol. Biol.* **2009**, *393*, 978−992.

(24) Kim, M. W.; Chelliah, Y.; Kim, S. W.; Otwinowski, Z.; Bezprozvanny, I. *Structure (Oxford, U. K.)* **2009**, *17*, 1205−1212.

(25) Walters, R. H.; Murphy, R. M. *J. Mol. Biol.* **2011**, *412*, 505−519.

(26) Chen, S.; Berthelier, V.; Yang, W.; Wetzel, R. *J. Mol. Biol.* **2001**, *311*, 173−182.

(27) Masino, L.; Kelly, G.; Leonard, K.; Trottier, Y.; Pastore, A. *FEBS Lett.* **2002**, *513*, 267−272.

(28) Wetzel, R. *J. Mol. Biol.* **2012**, *421*, 466−490.

(29) Bhattacharyya, A.; Thakur, A. K.; Chellgren, V. M.; Thiagarajan, G.; Williams, A. D.; Chellgren, B. W.; Creamer, T. P.; Wetzel, R. *J. Mol. Biol.* **2006**, *355*, 524−535.

(30) Perutz, M. F.; Finch, J. T.; Berriman, J.; Lesk, A. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5591−5595.

(31) Miettinen, M. S.; Knecht, V.; Monticelli, L.; Ignatova, Z. *J. Phys. Chem. B* **2012**, *116*, 10259−10265.

(32) Landrum, E.; Wetzel, R. *J. Biol. Chem.* **2014**, *289*, 10254−10260.

(33) Ferrone, F. A. *J. Mol. Biol.* **2015**, *427*, 287−290.

(34) Crick, S. L.; Ruff, K. M.; Garai, K.; Frieden, C.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 20075−20080.

(35) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773−778.

(36) Marchut, A. J.; Hall, C. K. *Proteins: Struct., Funct., Genet.* **2007**, *66*, 96−109.

(37) Thakur, A. K.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 17014−17019.

(38) Lapidus, L. J.; Steinbach, P. J.; Eaton, W. A.; Szabo, A.; Hofrichter, J. *J. Phys. Chem. B* **2002**, *106*, 11628−11640.

(39) Olshina, M. A.; Angley, L. M.; Ramdzan, Y. M.; Tang, J.; Bailey, M. F.; Hill, A. F.; Hatters, D. M. *J. Biol. Chem.* **2010**, *285*, 21807−21816.

(40) Legleiter, J.; Mitchell, E.; Lotz, G. P.; Sapp, E.; Ng, C.; DiFiglia, M.; Thompson, L. M.; Muchowski, P. J. *J. Biol. Chem.* **2010**, *285*, 14777−14790.

(41) Morozova, O. A.; Gupta, S.; Colby, D. W. *FEBS Lett.* **2015**, *589*, 1897−1903.

(42) Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P. *Science (Washington, DC, U. S.)* **2012**, *338*, 1177−1183.