JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

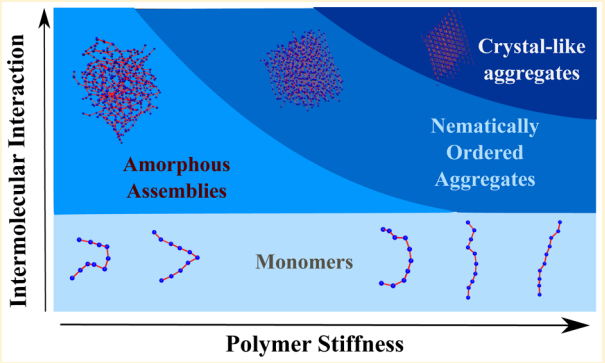
Defining a Physical Basis for Diversity in Protein Self-Assemblies Using a Minimal Model

Srivastav Ranganathan,* Samir K. Maji,* and Ranjith Padinhateeri*

Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Mumbai 400076, India

Supporting Information

ABSTRACT: Self-assembly of proteins into ordered, fibrillar structures is a commonly observed theme in biology. It has been observed that diverse set of proteins (e.g., alpha-synuclein, insulin, TATA-box binding protein, Sup35, p53), independent of their sequence, native structure, or function could self-assemble into highly ordered structures known as amyloids. What are the crucial features underlying amyloidogenesis that make it so generic? Using coarse-grained simulations of peptide self-assembly, we argue that variation in two physical parameters—bending stiffness of the polypeptide and strength of intermolecular interactions—can give rise to many of the structural features typically associated with amyloid self-assembly. We show that the interplay between these two factors gives rise to a rich phase diagram displaying high diversity in aggregated states. For certain parameters, we find a bimodal distribution for the order parameter implying the coexistence of ordered and disordered aggregates. Our findings may explain the experimentally observed variability including the "off-pathway" aggregated structures. Further, we demonstrate that sequence-dependence and protein-specific signatures could be mapped to our coarse-grained framework to study self-assembly behavior of realistic systems such as the STVIIE peptide and Aβ42. The work also provides certain guiding principles that could be used to design novel peptides with desired self-assembly properties, by tuning a few physical parameters.



INTRODUCTION

Protein self-assembly is associated with a wide range of native functions in biology.[1−7] Additionally, aberrant self-assembly of proteins/peptides into ordered β-sheet rich fibrillar entities known as amyloids has been implicated as the causative factor in various pathophysiological conditions including Alzheimers' and Parkinsons' diseases.[1,8,9] Structurally, these fibrils are composed of repeating subunits of β-strands contributed by the constituent proteins/peptides. These β-strands are stacked in parallel or antiparallel orientation, perpendicular to the axis of the fibril.[10] Owing to their β-sheet rich nature, amyloids possess a high degree of order resulting from the regular, linear array-like arrangement within the filament. However, protein aggregates are not always ordered in nature.[1,8,11,12] Natively unstructured peptide/protein chains or those that are incorrectly folded could sometimes assume disordered aggregated forms, which may be stable or a metastable state en route to forming linearly ordered structures.[13−15]

The examples of proteins/peptides that have the ability to self-assemble into ordered structures (like amyloids) are highly diverse with respect to their sequence, native structure or functions.[1,2] Interestingly, previous studies have also shown that even those proteins/peptides typically not associated with amyloid formation could be driven to form amyloid-like structures by suitably altering the solution conditions.[16,17] The diversity in amyloidogenic protein/peptides suggests that the ability to self-assemble could emerge out of certain generic properties (beyond sequence, native structure, etc.) of the polypeptide chains.[1,2,18−20]

The remarkable similarity in the higher-order structural signatures (e.g., the β-sheet rich nature) despite the large diversity in the proteins/peptides that constitute the "amylome"—the universe of all peptides that can form amyloids—is paradoxical.[21] This leads to the hypothesis that it could be the physical properties of the polypeptide chain and not the specific molecular details that contribute to this widely observed commonality in amyloids. Experimental findings that establish parallels between amyloid fibrils and nematic liquid crystals suggest that the geometric properties of the polypeptide chain could drive their assembly into the ordered "structural phases".[22,23] One of these geometric properties that shows a strong correlation to amyloid formation is the tendency of the polypeptide chain to assume extended β-strand like structures.[24−26] Furthermore, molecular dynamics simulation studies also reveal that the amyloidogenic regions in proteins are less flexible in comparison to their nonamyloidogenic counterparts.[27] These findings make this phenomenon apt for a generic physical polymer description. This gives rise to the following question—what are the signatures of protein

Received: June 22, 2016
Published: October 3, 2016

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

aggregation that could be attributed to the intrinsic physical properties of the self-assembling polypeptide chain? Answering this question would enable us to provide a broad physical description of the phenomenon of protein aggregation and amyloid formation.

In the current study, we provide a physical basis for the diversity in protein self-assemblies and the tendency of peptides/proteins to self-assemble into ordered aggregates. The study attempts to identify signatures of protein aggregation that arise out of the fundamental physical properties of the polymer chain such as self-interaction propensity and backbone flexibility. We begin with a minimal self-assembling homopolymeric system and further extend our findings for heteropolymeric assemblies. To model the self-assembly phenomenon, we employ a coarse-grained semiflexible polymer model with only short-range interactions and excluded volume effects. Using Brownian dynamics simulations, we delineate the role played by polymer flexibility and interaction strengths on the diversity of self-assembled structures, in the presence of thermal fluctuations. By systematically varying these two parameters, we identify various aggregate geometries that could be relevant to biological self-assemblies. From the discontinuous disorder−order transition and bimodality in order parameter near the transition point, we provide a plausible explanation for the variability in self-assembled states and experimentally observed "off-pathway" aggregates. Our findings suggest that the presence of an "interaction patch" (within a polymer) plays a vital role in governing their aggregate shape. We further provide a rationale for the sequence-dependence of amyloid formation. Also, we extend the findings to demonstrate the applicability of the simple model in capturing well-characterized structural signatures in Aβ42 amyloid fibrils. Overall, the current study demonstrates how the interplay between polymer flexibility, interaction strength and thermal forces could influence the phenomenon of protein aggregation.

METHODS

Molecular simulations are powerful tools to study secondary structural transitions in proteins/peptides during aggregation.[28−30] Atomistic simulations aid in the structural understanding of monomers and small-oligomeric entities, providing us with vital insights into the early stages of protein aggregation[31] and also help us probe fibril stability.[32,33] However, the applicability of atomistic simulations is limited severely by the time-scales and length-scales they can access, thereby necessitating the need for coarse-grained models. Discrete models have been used to understand the relationship between molecular structure and aggregate morphology[34] and the effect of β-sheet propensity on aggregation.[26] In general, these models are efficient tools to understand how polymer systems could undergo transition between various aggregated states.[35]

In order to study the diversity in aggregated structures, we performed Langevin dynamics simulations. In our study, we consider N semiflexible polymer chains, each made up of M beads, in a cubic box with periodic boundary conditions. In all our simulations, the number density of beads, $\phi = M{\cdot}N/V$, is kept constant, where V refers to the volume of the cubic box. The chains in the box have the following interactions. The neighboring beads in a polymer chain are bonded via the harmonic potential with energy

$$E_{\text{stretching}} = k_s \sum_{i=1}^{M-1} (|\vec{r}_i - \vec{r}_{i+1}| - r_0)^2 \tag{1}$$

where $\vec{r}_i$ and $\vec{r}_{i+1}$ refer to i^{th} and $(i+1)^{\text{th}}$ bead positions, respectively; r_0 refers to the equilibrium bond length and k_s represents the spring constant. This interaction ensures the connectivity between the beads of a polymer chain. To model bending rigidity, any two neighboring bonds in a polymer interact via the following potential

$$E_{\text{bending}} = \kappa \sum_{i=1}^{M-2} (1 - \cos\theta_i) \tag{2}$$

where θ_i refers to the angle between i^{th} and $(i+1)^{\text{th}}$ bond, and κ is the bending stiffness. All other nonbonded, interbead interactions were modeled using the Lennard-Jones (LJ) potential,

$$E_{\text{nb}} = 4\epsilon \sum_{i<j} \left[\left(\frac{\sigma}{|\vec{r}_i - \vec{r}_j|} \right)^{12} - \left(\frac{\sigma}{|\vec{r}_i - \vec{r}_j|} \right)^{6} \right] \tag{3}$$

for all $|\vec{r}_i - \vec{r}_j| < r_c$, where r_c refers to the cutoff distance beyond which the nonbonded potentials are neglected. The cutoff for the LJ calculations were kept at 2.5 times of σ. Note that this function has two parts: a part that models steric repulsion when the beads overlap and an attractive part otherwise. For ease of implementation, in this model, this attractive part is used as an *effective interaction* that accounts for all short-range attractive forces. ϵ signifies the strength of the attractive interaction, and has the units of energy. Note that the bending energy parameter (κ) controls the stiffness of the individual polymer chains while the LJ potential parameter (ϵ) controls the interaction between any two beads (interchain and intrachain).

Simulation Parameters. In our study, we used the LAMMPS molecular dynamics package to perform the dynamic simulations,[36] where the simulator solves Newton's equations with viscous force, and a Langevin thermostat ensuring an NVT ensemble with temperature T = 310 K. An integration time step (dt) of 15 fs was used for the simulations. The mass of each bead was considered to be 110 Da (average mass of amino acids). The size parameter for the beads is taken as σ = 4.5 Å. The parameters for the bonded springs was fixed as r_0 = 4.4 Å and k_s = 10 kcal/mol. The damping time for the Langevin thermostat was 1.2 ps. Similar values for these parameters have been previously used by Bellesia et al. to perform coarse-grained protein simulations.[26]

Choice of ϵ and κ Values and the Rationale. In this study, we systematically vary the parameters ϵ and κ to investigate self-assembly. The strength of interaction, ϵ, models the *effective interaction* between any two amino-acid beads, including H-bonds, hydrophobic effect and van der Waals interaction. The range of ϵ in our study varies from 0.25 to 1 kcal/mol, which is of the same order of magnitude as that was reported previously.[37,38] For example, the ϵ values for interaction between residue beads in the MARTINI force field, a commonly used coarse-grained protein model, also ranges from 0.4 to 1.2 kcal/mol.[37] It has been previously shown that the strength of individual hydrogen bonds that stabilize the NNQQNY fibrils is in the range of 1.2 to 2.9 kcal/mol.[38] The range we chose in our study was not only comparable to the interaction strengths reported in protein literature but also suitable for studying the interplay between flexibility and thermal fluctuations. Further, we chose the κ values based on previous estimates of persistence length (L_p) of polypeptide chains and segments within larger proteins.[39−41] The distribution of persistence length of stretches within proteins/peptides[39] suggests that the L_p can range from 15 to 60 Å, which is about 5 to 20 amino-acid residues in length. The values for bending stiffness (κ) used in this study fall in the range of 2.5 to 10 kcal/mol, which is equivalently a persistence length in the range of 5 to 20 beads in length. The parameter values chosen in this study are thus relevant for probing protein self-assembly.

Order Parameter. One of the key objectives of the work is to study the effect of varying bending stiffness and interaction strength on the nature of the aggregates, as characterized by their geometries. In order to characterize the orientational order in the aggregate, we used the nematic order parameter (S), which is a common measure of the extent of orientational ordering in liquid crystals.[42] Ceccini and Rao have previously used the orientational order parameter to monitor the amyloid formation by Aβ peptide using atomistic molecular dynamics simulations.[43,44] The nematic order parameter S is given by,

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

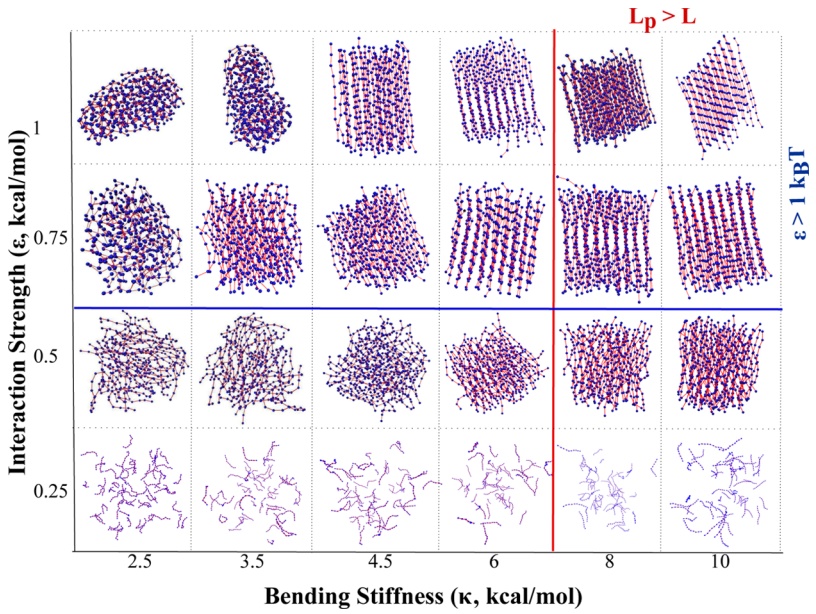


Figure 1. The $\kappa-\epsilon$ phase-space. Aggregate morphology in response to varying bending stiffness and interaction strengths, represented by the equilibrium structures from the Langevin dynamics simulations. The phase diagram shows that a critical interaction strength is required for aggregation to occur. Polymers with lower bending stiffness form assemblies with globular symmetry. On the other hand, stiffer polymers form aggregates with higher order. Ordered aggregates at lower interaction strengths could have varying internal arrangement in comparison to higher ϵ. As we move from the lower (0.5 kcal/mol) to higher (1 kcal/mol) interaction strengths, the structures approach crystal-like arrangement. The red vertical line represents the point beyond which $L_p > L$ while the blue horizontal line represents the point beyond which $\epsilon > 1\ k_BT$. The simulations were performed using 60 polymer chains that are 10 beads long, at a concentration of ≈30 mM.

$$S = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}(\hat{z}_i\cdot\hat{d})^2 - \frac{1}{2} \tag{4}$$

where $\hat{d}$ is the director vector that describes the average direction of the alignment. $\hat{z}_i$ is the molecular vector that describes any polymer in the system. In this study, we define $\hat{z}_i$ as the vector connecting the two terminal beads of i^{th} polymer chain. S can take any value from 0 to 1. $S \rightarrow 1$ suggests a highly ordered system while $S \rightarrow 0$ suggests an entirely disordered system.

Further, to identify differences in local and long-range order between various structures with similar nematic order, we used the radial pair-distribution function (RDF) denoted as $g(r)$ and the static structure factor, $S(q)$. The radial distribution function is a commonly used quantity while studying arrangement of molecules in a self-assembly or a crystal.[28,45] In our study, the $g(r)$ describes the average density correlations between particles at any given distance r, normalized to the corresponding density for an ideal gas. We used the $g(r)$ plugin in VMD,[46] with suitable modifications, to compute the distribution functions for our simulations. The static structure factor, $S(q)$ is the Fourier space equivalent of $g(r)$, which would give us information accessible in scattering experiments.[47−49] In this study, we used $S(q)$ to identify differences in long-range ordering between various structures.

We performed the simulations for a time (≥4.5 μs) that is long enough for the system to reach steady state having constant mean energy and order parameter (S). We ensured that the structures are in equilibrium by starting from very different initial configurations and confirmed that they lead to the same final state. All structures presented here were rendered using VMD molecular visualization program.[50]

RESULTS

In this section, we present results from our simulations with N interacting semiflexible polymers, each of length L, in a cubic box with periodic boundary conditions as described in the Methods section. Each polymer chain in the simulation box represents short peptides that can self-assemble into diverse aggregated structures. Short peptides or segments within a large protein have been shown to possess the ability to form amyloid-like structures.[51−53] First, we study the self-assembly of short homopolymer chains (contour length $L \leq 10$ beads long) to identify the fundamental factors governing their self-assembly. Studies have shown the ability of homopolymeric peptides like polyalanine and polyglutamine to self-assemble into fibrils.[51−53] We then extend the findings from these homopolymer simulations to heteropolymeric aggregation.

Aggregate Diversity: Interplay between Polymer Flexibility and Interaction Strength. The aim of our study is to identify the physical properties of peptide chains that dictate their assembly into either ordered amyloid-like structures or into globular aggregates with negligible order. To further this objective, we focus on two parameters: the flexibility of the constituent polymer chain (κ) and the strength of interaction between any two monomeric beads (ϵ). First, we simulated $N = 60$ peptide chains, each of which are 10 beads long, as described in the Methods section, at a concentration of ≈30 mM. Upon varying the values of the parameters κ and ϵ, we obtained different aggregated phases as shown in Figure 1. For very low interaction strengths (ϵ = 0.25 kcal/mol), we observe an isotropic state with the polymer chains scattered in solution (Figure 1 lowest row). However, for larger values of ϵ, we find self-assembled macro-structures with diverse morphologies. At these higher interaction strengths (Figure 1, $\epsilon > 0.25$ kcal/mol), as the constituent polymer chains become stiffer (increasing κ), the aggregates display increased ordering, with the individual chains aligning themselves within the aggregate. In other words, flexible polymer chains self-assemble into aggregates with globular symmetry. Stiffer polymers, on the other hand, form structures with the constituent chains aligned on an average along an axis (see Figure 1). For high κ and ϵ, the polymer chains self-assemble into a stack of sheets with varying

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

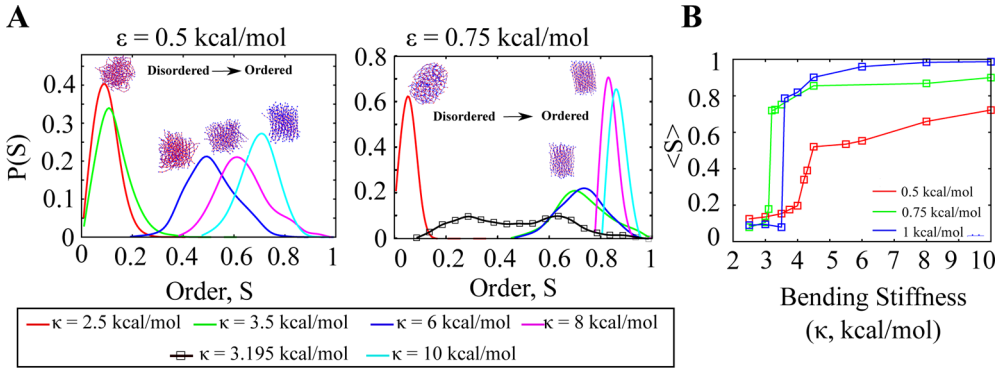


Figure 2. The disorder−order transition. (A) Aggregated states accessed by the polymer chains for ϵ = 0.5 kcal/mol (left panel) and 0.75 kcal/mol (right panel) characterized by the distribution of orientation order parameter, $P(S)$. The bimodal distribution (black curve, ϵ = 0.75 kcal/mol) shows the coexistence of ordered and disordered states near the transition point. (B) Transition from the disordered to ordered state in response to varying bending stiffness at different interaction strengths (red, green and blue refer to ϵ values of 0.5, 0.75, and 1 kcal/mol, respectively).

degrees of order (or disorder) depending upon their location on the κ−ϵ phase-space (Figure 1). As seen in Supporting Information Figure S1 and the Supporting Information Videos S1, S2 and S3, the ordered structures formed by stiff polymer chains (κ > 8 kcal/mol) manifest varying degrees of order depending on the interaction strength. Overall, the inherent stiffness that is critical for ordering would depend on a three-way competition among peptide flexibility, intermolecular interactions and thermal fluctuations. Therefore, for certain parameter regimes, even flexible chains can self-assemble into ordered structures (ordered structures to the left of $L_p > L$ line in Figure 1). Please note that the range of parameter values used here are relevant to protein interactions (see Methods section).

The concentration of polymer chains in the box in Figure 1 was approximately 30 mM. The concentration was chosen such that the peptides form a single aggregated cluster, enabling us to characterize the structural features of the aggregate in greater detail. However, at lower concentrations, the aggregates are smaller in size, suggesting that cluster sizes are a function of concentration and interaction strength (Supporting Information Figure S2 and S3). For any given value of ϵ, there exists a critical concentration for self-assembly to occur, with an increase in ϵ resulting in aggregation being favored at lower threshold concentrations.

To quantify the extent of ordering (or lack of it) within the aggregate, we computed the orientational order parameter S (see Methods section) for the self-assembled structures and plotted its distribution (Figure 2A). The distribution allows us to monitor the aggregated states sampled by the self-assembling polymer chains. S measures the extent of alignment of individual polymer chains along a reference axis. Here, we refer to disordered aggregates as those with $S \leq 0.5$ and ordered aggregates as those with $S > 0.5$. In both subpanels of Figure 2A, we vary the bending stiffness such that the persistence length (L_p) of the constituent polymer chain ranges from $L_p = 0.5L$ (flexible) to $L_p = 2L$ (stiff). L_p and L refer to the persistence length and the contour length (in number of beads), respectively. For an interaction strength of ϵ = 0.5 kcal/mol (1 k_BT), we observe that the aggregates constituted by the flexible polymer chains (κ < 6 kcal/mol) are predominantly disordered in nature, as evident from the distribution of the order parameter, which peaks at very low values of S (Figure 2A, ϵ = 0.5 kcal/mol, red and green curves). As the persistence length of the constituent polymer increases, the distributions shift rightward indicating an increase in the orientational order. Interestingly, when the L_p is roughly equal to the polymer length (L), the system accesses both ordered and disordered configurations, with a symmetric distribution peaking around S = 0.5 (Figure 2A, ϵ = 0.5 kcal/mol, blue curve). For values of κ where the contour length (L) is less than the persistence length (L_p), the aggregate prefers to access the ordered states. As the interaction strength increases such that ϵ > 1 k_BT, even flexible polymers ($L_p < L$) self-assemble into predominantly ordered structures (Figure 2A, right panel). In other words, the interaction between the chains is strong enough to hold the polymers in an extended configuration, allowing the system to access the ordered states even with flexible constituent chains (Figure 2A, ϵ = 0.75 kcal/mol, black and green curves). A similar trend is evident in case of an interaction strength of 1 kcal/mol, as seen in Supporting Information Figure S4.

The transition from the disordered to ordered state is of great biological significance. Previous studies suggest that during the process of protein/peptide self-assembly, various transient states could be encountered.[54,55] We thus study the transition from the disordered state to the orientationally ordered state with increasing κ. Our results highlight interesting differences in the nature of this transition at various values of ϵ. While we observe a sharp-transition from the globular aggregated state ($\langle S \rangle$ < 0.5) to the ordered state at ϵ > 0.5 kcal/mol, the transition is smoother at 0.5 kcal/mol (Figure 2B). The discontinuous nature of the transition is further highlighted by the bimodal distribution of the order parameter near the transition point (Figure 2A, black curve for ϵ = 0.75 kcal/mol). This bimodal signature confirms the coexistence of ordered and disordered structures in this intermediate regime. These findings suggest that in this regime of κ and ϵ, the polymer chains could display high variability, sampling both ordered and disordered states. Therefore, slight perturbations in the environmental conditions in this intermediate regime might result in dramatically different end-structures.

Polymer Chains Show Distinct Arrangement within Ordered Self-Assemblies. Diversity in aggregates is not restricted to differences between the amorphous state and the crystal-like ordered states. In fact, ordered filamentous aggregates have also been known to exhibit polymorphism.[56,57] While some of these structures, for example the β-barrel like state, exhibit toxicity, the others are nontoxic.[12]

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

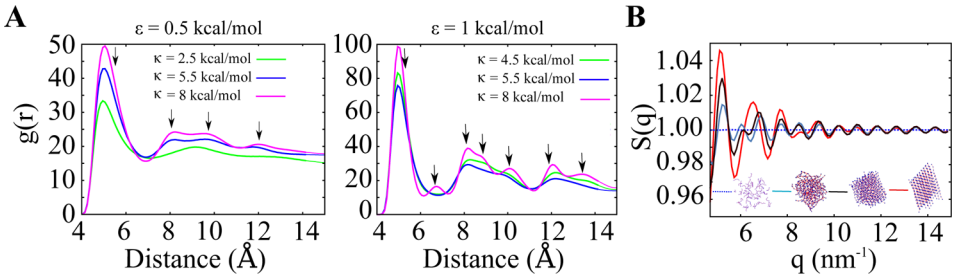


Figure 3. (A) The radial pair distribution function used to distinguish aggregate geometries. Peaks at longer length-scales indicate long-range order and thereby increased ordering in the system. Low degree of ordering in flexible polymeric systems is evident from very diffuse long-range peaks (green curve, left panel). Sharper long-range peaks emerge for higher interaction strength of 1 kcal/mol (right panel) in comparison to lower ϵ of 0.5 kcal/mol (left panel). (B) The static structure factor $S(q)$ reveals that the representative structures have similar short-range order but distinct long-range ordering.

In this context, the $\kappa-\epsilon$ phase-space shows that an interplay between the two parameters could result in a variety of different structures, both ordered and disordered. A natural progression is to ask the question whether all the ordered states ($S > 0.5$) possess similar underlying arrangement of constituent chains? A way to quantify the variation in internal packaging of polymer chains within such assemblies is to compute quantities like radial distribution function ($g(r)$) and the static structure factor ($S(q)$), which can identify short and long-range order within these aggregates. For disordered globular aggregates, we observe no correlation for length-scales >4.8 Å (Figure 3A, ϵ = 0.5 kcal/mol, green curve). A shift from the globular aggregates to a more ordered state with an increase in bending stiffness is evident from an emergence of long-range peaks in the $g(r)$ plot (Figure 3A, ϵ = 0.5 kcal/mol, see arrows). As we strengthen the interactions (Figure 3A, ϵ = 1 kcal/mol), there is an increase in both the number and the sharpness of the long-range peaks: a clear measure of differences in the way the structures get packaged at higher ϵ. For instance, while extremely stiff polymer chains (κ = 10 kcal/mol) associate into a stack of sheets at ϵ = 1 kcal/mol, the assemblies show much lower long-range order at smaller values of ϵ (see Figure 1 and Supporting Information Videos S1, S2 and S3). This shift manifests itself via a change in peak position and sharpness in the $g(r)$ plot. The structure factor $S(q)$ further highlights these underlying structural differences (Figure 3B). As evident from the figure, the $S(q)$ for predominantly isotropic system at ϵ = 0.25 kcal/mol (blue dashed curve) is flat. As self-assembled structures appear, peaks start to emerge. Interestingly, the two ordered structures (black and red curves in Figure 3B) differ significantly in the long-range peak positions (at small values of q) and their amplitudes. The highly ordered assembly of sheets (red curve in Figure 3B) displays a great degree of periodicity, indicating a crystal-like arrangement of the polymer chains within the structure. The difference in the peak positions and their amplitudes among the different ordered structures confirms presence of structural diversity. Thus, the underlying packaging of ordered aggregates is also an interplay of κ and ϵ.

The Length of the "Interaction Patch" Is a Key Factor in Ordered Self-Assembly. Our results so far establish the role of stiffness in promoting ordered aggregation. A key question that arises out of this observation is whether all stiff self-interacting polymers would assemble into ordered aggregates, regardless of the length of the constituent chain. To probe the role of length on the nature of the aggregate, we perform simulations by systematically varying two parameters, the chain length (L) and interaction strength (ϵ), keeping bending stiffness constant at κ = 4 kcal/mol (equivalently, L_p = 6 beads). The number density of the interacting beads, $\phi = M \cdot N/V$, in the simulation box was maintained constant for all the simulations in order to ensure that the volume fraction of the interacting chains is the same. Therefore, the results of length dependence reported below are not due to differences in volume fraction (previously known to affect the order−disorder transition[58−60]). As seen in Figure 4, for short polymer chains

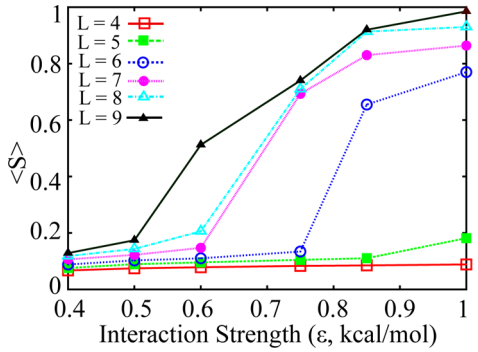


Figure 4. Effect of polymer length on ordering. Average order in the aggregate as a function of interaction strength, for polymeric chains of varying lengths, at a fixed κ of 4 kcal/mol. The interaction strength at which we observe ordered structures is higher for shorter polymer chains. The volume fraction of chains in the box was maintained constant for all the simulations.

(L = 4 or L = 5), we see negligible order (low value of $\langle S \rangle$) within the aggregate. This is despite the chain length (L) being shorter than the persistence length. For large values of ϵ, as we go from long to shorter polymers, we see a decrease in the orientational order (also see Supporting Information Figure S6). The lack of ordering at lower lengths could be a result of increased mobility for smaller polymers as mobility is inversely proportional to the length. Note that the aspect ratio (length/diameter) for these small chains (L = 4 and L = 5) in these examples, is typically small and that might also lead to disfavoring of order.[61]

Therefore, for a homopolymer of any given length L, there exists a critical interaction strength beyond which one would observe ordering. Shorter polymer chains would require stronger interactions to assemble into a nematic-like state (Figure 4 and Supporting Information Figure S6). Our results clearly suggest that aggregate shape is not a function of polymer stiffness alone. A critical length of the interacting residues or an

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

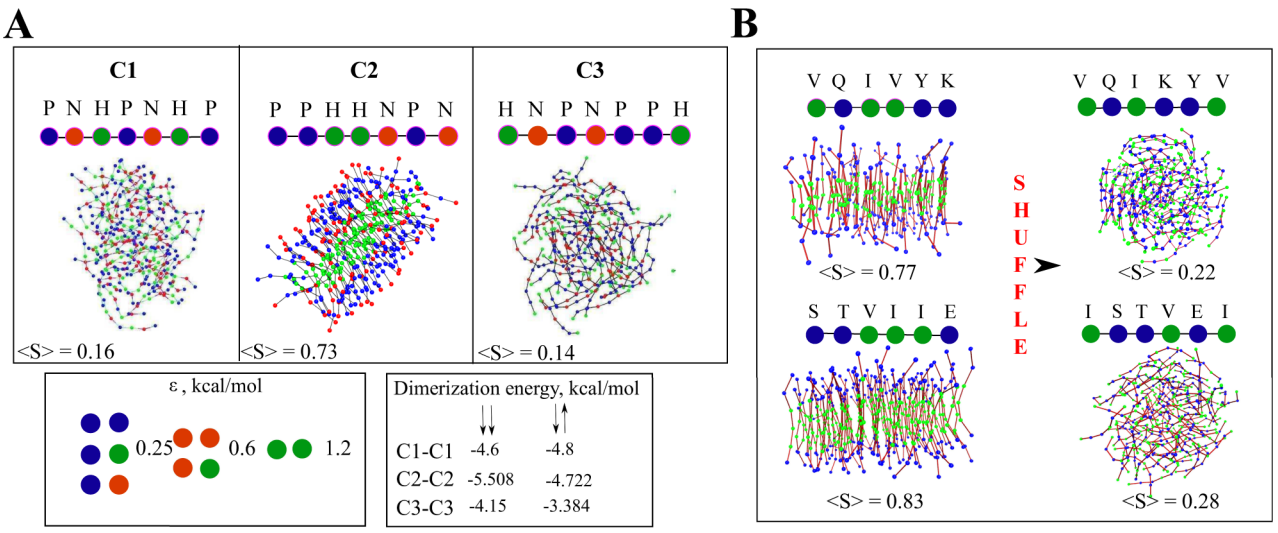


Figure 5. Role of sequence on aggregate geometry. Simulations of polymer chains with three types of beads, P (blue), N (red) and H (green), which are analogous to polar, slightly hydrophobic and extremely hydrophobic residues, respectively. (A) Three sequences (labeled C1, C2 and C3) with the same combination of beads but different arrangement of residues were used to perform self-assembly simulations. Only one of the three polymer chains (with sequence C2) assembles into ordered structures. The corresponding dimerization energies show that C2 is the most stable of the three sequences in its extended dimeric form. The pairwise interaction strengths for the unique bead-pairs are shown in the lower panel. (B) Simulations using known amyloidogenic sequences (VQIVYK and STVIIE) with hydrophobic residues mapping to the H-bead type (green beads) and polar residues mapping to the P-bead type (blue beads) were performed. The sequences VQIVYK and STVIIE assemble into ordered aggregates. Shuffling of the sequences leads to loss of order within the aggregate. The pairwise interactions used here are the same as in subpanel A.

"interaction patch" might thus be essential to form ordered aggregates.

Role of Sequence in Aggregate Geometry. The results discussed so far elucidate the role of polymer flexibility, length and the interaction strength in governing the nature of homopolymeric aggregates. However, protein aggregates are seldom constituted of homopolymeric chains; rather the propensity to aggregate as well as the nature of the aggregated structure both display amino-acid sequence dependence. Previous studies have established the importance of amino-acid sequence on conformational dynamics[62,63] and amyloidogenic propensity, emphasizing that fibril formation is not dictated by amino-acid composition alone.[21] Using our model, we attempt to understand whether sequence dependence could partially arise due to differences in interaction patterns within the aggregated state. In our simulations we assume that the constituent amino-acid beads are capable of forming H-bonded interactions and spatially compatible within the β-sheet geometry. We introduced sequence dependence into the model by constructing polymer chains with three different constituent bead types P, N and H that mimic polar, slightly hydrophobic and highly hydrophobic amino acid beads, respectively. Each unique pair of beads can interact with varying strengths based upon their physicochemical property. For simplicity, we assume that the strength of association of polar residues is the lowest, while the hydrophobic residues have the highest affinity toward each other (see Figure 5A, bottom left panel). Using these three bead types, we construct three different polymer chains that vary in sequence but possess the same overall composition of monomer beads (see Figure 5A, top panel). We performed self-assembly simulations for the three polymer chains (labeled C1, C2 and C3 in the Figure 5A) to test whether sequence-dependence of the aggregated state could arise from differential interaction strengths alone. Interestingly, we observe that only one of the three chains (C2) self-assembles into an ordered structure ($\langle S \rangle > 0.5$) while the other two assume globular morphologies ($\langle S \rangle < 0.5$). To further probe the basis of this sequence-dependence of ordered self-assembly, we estimate the compatibility of the constituent chains to assume an extended geometry during self-assembly. To do so, we force the three polymer chains (C1, C2 and C3) to assume aligned parallel ($\downarrow\downarrow$) and antiparallel ($\downarrow\uparrow$) dimeric structures and estimate the net interaction strength of the dimeric structures (Figure 5A, bottom right panel). Interestingly, the sequence C2 that assembles into an ordered aggregate also has the highest overall strength of interaction in its parallel orientation. The strength of the C2 dimers was found to be 5.5 kcal/mol, which is much more stable in comparison to the other dimeric states (C1−C1 and C3−C3) whose strengths were <5 kcal/mol. These results suggest that C1 and C3 are less compatible for self-assembling into extended, aligned structures with nematic order. The assemblies constituted by these chains thus favor globular structures.

We further use the model to study the effect of sequence shuffling on two peptides, VQIVYK[64] and STVIIE,[65] that have been reported to form amyloids in vitro. VQIVYK, is an aggregation-prone segment from the microtubule-associated tau protein[64] whose sequence is composed of 3 highly hydrophobic residues (2 valines and 1 isoleucine) and 3 polar residues (glutamine, tyrosine and lysine). In the native sequence VQIVYK, the three hydrophobic residues are in close proximity to each other. To test whether the arrangement of polar and hydrophobic residues along the sequence could result in altered aggregated structures, we constructed a sequence VQIKYV where the hydrophobic residues are scattered across the chain. Using the relative interaction affinities described in Figure 5A, we performed simulations to probe the effect of shuffling on aggregate order (Figure 5B). As shown in Figure 5B, the native sequence VQIVYK self-assembles into an aggregate with high orientation order ($\langle S \rangle$ = 0.77). However, the assembly formed by the shuffled sequence VQIKYV possesses negligible order ($\langle S \rangle$ = 0.22) despite the

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

same overall amino-acid composition. This loss of ordered aggregation ability upon sequence rearrangement is in agreement with experiments by Goldschmidt et al., where the shuffling of VQIVYK has been shown to result in diminished amyloidogenic propensity.[21]

To further demonstrate the efficacy of our model in capturing sequence effects, we also shuffled another known amyloid-forming sequence STVIIE, which has previously been used to determine the position specificity of amino acids within a sequence.[65] Positional scanning mutation studies have previously revealed that the third, fourth and fifth positions of the peptide are the most restrictive for amyloid formation. Any alterations to these positions was thus reported to result in a loss of amyloidogenic activity.[65] Using this rationale, we designed a sequence ISTVEI where the three core positions were disrupted. The self-assembly simulations of the STVIIE peptide reveal that the native sequence can self-associate to form nematically ordered structures (Figure 5B). However, the aggregate formed by the shuffled sequence ISTVEI shows a drastic loss of order, thereby confirming the experimental findings.[65]

Further, we tested another mutant of the STVIIE peptide, STLNFE, which has a lower hydrophobicity in the core stretch (isoleucine replaced by glutamine). When we performed simulations using the STLNFE sequence (Supporting Information Figure S7), we observe small clusters of aggregated chains with negligible order, unlike the highly ordered structures formed by STVIIE. This is also consistent with experimental findings showing that while STVIIE can undergo fibril formation upon incubation, negligible self-assembly was observed for the mutant form STLNFE.[65]

Overall, we predict that a shuffling of the sequence could lead to a dramatic loss of order, suggesting that the precise network of interactions plays a key role in dictating aggregate geometry. The simulations also suggest that a stretch of residues with high self-interacting propensity in close vicinity could favor assembly into ordered structures. This is further corroborated by the fact that amyloidogenic sequences are typically rich in hydrophobic residues or residues that allow charge-based stabilization (salt bridges) of the aggregated state.[66]

Destabilization of Interactions Could Result in Varying End-Structures. Biopolymer self-assemblies are dynamic entities that respond to factors that modulate their interaction strengths.[67−70] It is thus important to understand whether the stable assemblies that form under a given condition (in this study, a set of parameter values, ϵ and κ) respond to changes in interaction parameters. Here, we start with a preformed ordered aggregate (polymer chain C2 in Figure 5A) and mimic changes in solution conditions by changing the interaction strength, ϵ. This is based on the assumption that effect of any external factor (pH, salt concentration etc.) could be mapped onto an effective interaction strength ϵ. In this context, we probe the response of the aggregated state to destabilization of various interactions (Figure 6). In "Case 1", we weaken the interactions involving the N bead type (red beads) by changing the ϵ from 0.6 to 0.3 kcal/mol. The structures that ensue show a loss in orientational order as a result of the weakened interactions. However, in "Case 2", where the interactions involving the H bead type were weakened, one sees a complete disintegration of the aggregated state. Therefore, any attempt to destabilize the aggregate could lead to either the restructuring of the aggregated state or a complete disassembly of the supramolecular structures. A mere

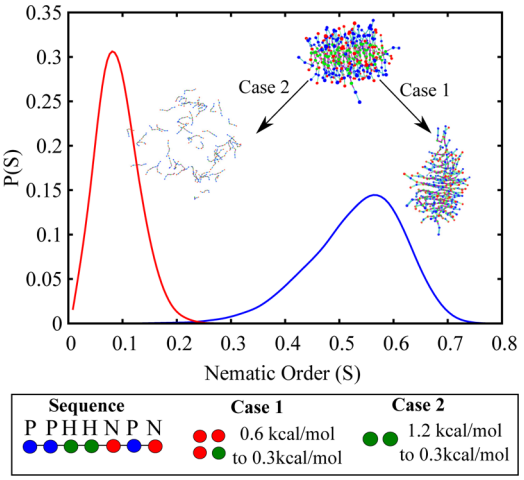


Figure 6. Destabilizing the aggregate. Starting with preformed ordered aggregates for the sequence C2 from Figure 5A, we see the effect of selectively destabilizing the interactions on the state of the system. Case 1 refers to destabilizing interactions involving the N beads while Case 2 involves destabilizing the H beads. The two scenarios, "Case 1" and "Case 2" result in loss of order (distribution peaking around 0.5) and complete disassembly (peak of the distribution around 0.1), respectively.

destabilization of the aggregate in a nonspecific manner might thus result in structures that are more deleterious than the stable fibrillar state.

Toy Model for Aβ42 Aggregation. Protein sequences typically possess patches with high β-sheet propensity scattered across the length of the chain. Also, segments with high self-interaction potency are also interspersed across the amino-acid sequence.[21] These patterns give rise to the distinct amyloid-formation tendencies for various proteins/peptides as well as their unique structural signatures. With an understanding of the necessary conditions for ordered aggregation, we ask the question whether the introduction of heterogeneous patches of varying self-interaction strength and stiffness along the length of a polymer chain could give rise to structural signatures observed in well characterized fibrils. We choose Aβ42, a peptide associated with Alzheimers disease as our model system. The structural details of amyloid fibrils formed by Aβ42 peptide have been characterized using quenched hydrogen/deuterium-exchange NMR studies.[71] Additionally, the TANGO aggregation algorithm[72] predicts the presence of two putative amyloidogenic regions with high β-aggregation propensities that lie between the regions 15−22 and 28−42 in the amino acid sequence. Using the existing knowledge of the Aβ42 peptide, we extend the findings from our coarse-grained description to develop a toy model for Aβ42 aggregation. Here, we account for varying self-interaction and β-sheet propensities along the peptide chain by considering a 42-bead long chain that is constituted by two kinds of monomer beads along its length: (i) the beads labeled R1 in Figure 7B refer to weakly interacting patches with high main chain flexibility and (ii) the beads labeled R2 in Figure 7B refer to regions with high hydrophobicity and β-sheet propensity. The corresponding interaction strengths for R1-R1, R1-R2 and R2-R2 were chosen to be 0.2, 0.2, and 0.8 kcal/mol, respectively. The choice of ϵ and the higher strength of interaction between the R2 beads, in comparison with the other two interaction pairs is driven by the hydrophobic nature of the β-aggregation core of the Aβ42 peptide. Therefore, in addition to forming the β-sheet

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

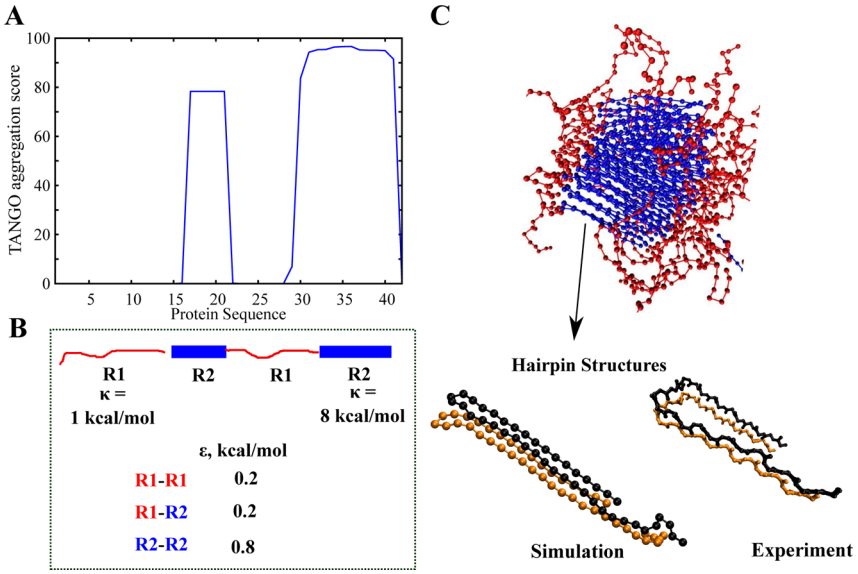


Figure 7. The Aβ42 toy model. (A) The putative regions with high β-sheet propensity according to the TANGO algorithm. Two regions, 14−22 and 26 to 42 score highly for β-aggregation propensity. (B) Mapping of the aggregation propensity along the Aβ42 sequence onto the toy model through our model parameters ϵ and κ. R1 and R2 refer to the regions with low and high β-aggregation propensity, respectively. (C) The self-assembled structures formed by polymer chains with two aggregation prone stiff patches (R2, blue beads) interspersed along the sequence with the more flexible regions (R1, red beads). The fundamental dimeric unit shows a hairpin like structure similar to the ones found in NMR experiments (PDB ID: 2BEG).

stabilizing hydrogen bonds, this region also favors self-interactions owing to its high hydrophobic content. The stronger interactions between the R2 beads is used to mimic the phenomenon of hydrophobic collapse in the Aβ42 peptide. This differential interaction is a standard method used in coarse-graining approaches for a wide range of problems: for example, to study assembly of lipid bilayers having hydrophobic/hydrophilic interactions.[73] Such an approach of using differential interaction strengths is also employed by the MARTINI coarse-grained protein force-field in order to simulate the hydrophobic effect.[37] The bending stiffness in the two regions R1 and R2 were assigned to be 1 and 8 kcal/mol, respectively. The higher order structures that were observed upon self-association show that the R2 regions (blue beads in Figure 7B) form the core of an ordered aggregate, flanked by the disordered, flexible R1 regions (red beads in Figure 7B). Interestingly, when we take a closer look at the organization of the individual polymer chains within the aggregate at the level of the dimer, we find that individual chains fold into hairpin-like structures that associate with other hairpin-like structures. This is in direct agreement with previous NMR studies that show that the hairpin-dimer, with each polypeptide chain contributing two self-interacting β-strands, is the most fundamental repeat structure within the protofilament.[71] The findings show that a coarse-grained, two-letter description of the peptide chain (by defining varying patches of κ and ϵ), can capture key structural signatures associated with known fibrils.

DISCUSSION

Applicability of a Minimal Physical Model. Self-assembly of proteins/peptides into supramolecular structures is a recurring feature in living systems. Ordered protein/peptide self-assemblies composed of a repeating substructure of β-sheet rich polypeptides have been routinely associated with neurodegenerative diseases like Alzheimers' and Parkinsons', as well as native biological functions like hormone storage.[1,2] A plethora of studies have revealed that the property to form aggregated structures is not restricted to any particular class of proteins/peptides.[1] A large number of proteins/peptides usually not associated with aggregation have also been reported to aggregate under suitable solution conditions. For example, pH, salt concentration, presence of cations or glycosaminoglycans in solution have all been reported to induce protein aggregation.[16,18,74,75] This raises an interesting question: can we identify the minimal factors that are essential to explain the generic aggregation tendencies of the otherwise diverse set of proteins/peptides that constitute the amylome?

In this paper, we argue that the ability to self-assemble could be an inherent *physical* property of the polypeptide chain. Considering this hypothesis, we attempt to identify the fundamental physical factors that govern the nature of the aggregated state. In amyloids, the globular symmetry of the assembly is broken due to the extended β-sheet rich conformation of the constituent peptide chain. In this extended sheet, the β-strands are stabilized by interstrand hydrogen bonded interaction and other short-range interactions. These self-interacting β-rich chains give rise to the characteristic linear order in amyloid fibrils. This generic feature makes it feasible to study the amyloid aggregation-phenomena using a coarse-grained approach considering two physical factors—the bending stiffness (κ) and the strength of the intermolecular interactions (ϵ). These two parameters implicitly account for sequence and environmental variables. For instance, κ, which is a physical property of the polymer chain, would depend on its constituent building blocks (say, the amino-acid sequence). In our study, we use κ to modulate the conformational landscape of the polymer chain, with lower values of κ resulting in the chain sampling more compact structures and higher values resulting in a greater tendency to sample the extended states. Therefore, the bending stiffness could be considered as a surrogate measure of the intrinsic β-sheet propensity of the

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

polypeptide chain. The second parameter under consideration in this study, ϵ refers to the strength of the intermonomer interactions. This could be a net effect of solution conditions that govern the strength of interaction or a sequence-effect dictated by the type of constituent residues. This property is thus analogous to the aggregation propensity of a polypeptide chain. In summary, sequence or environmental factors could be mapped onto these two parameters, allowing us to sample a large phase-space relevant to protein aggregation.

Rationale for Excluding Finer Molecular Details. One of the key questions of the current study is how much molecular detail is essential for capturing the aggregation behavior of proteins/peptides? Our aim was thus to probe whether aggregate diversity could emerge from a minimal, physical description of the problem. Therefore, we employ a minimal model with an effective set of parameters to answer these questions. If a model lacking specific molecular details (side-chain information, directionality of bonds and chirality) could reproduce commonly observed features of aggregation, it would enable us to identify the signatures of self-assembly that could be attributed to the fundamental physical factors alone. Upon resolving these questions, the finer details could be systematically introduced to understand how each chemical detail (chirality, bond directionality, etc.) alters the phase behavior. Interestingly, despite our model not accounting for many of these finer chemical details, the emergence of typical protein-aggregation signatures was still observed.

From Monomers to Amorphous Aggregates to Ordered Assemblies. Heterogeneity in aggregated structures is a widely observed phenomenon in biological systems. The aggregated state of a polypeptide is a function of both intrinsic and extrinsic factors.[1,2,16−18] Also, several diverse morphologies could be encountered en route to populating the stable fibrillar state.[12,76] Diversity in self-assembled structures not only lends itself to functional diversity but also has tremendous significance in the context of cytotoxicity.[8,9,12,77] Understanding the factors that govern the nature of the aggregated state is thus vital.

Our simulations show that by varying the two phase parameters (κ and ϵ in Figure 1), one could drive the polymer chains to access various states, including monomers (low values of ϵ), disordered amorphous aggregates (low κ and low ϵ) and ordered aggregates (high κ). These findings allow us to answer the following question: starting from a highly rigid polymer chain, if we systematically reduce the stiffness (introduce semiflexibility), at what point would it result in a loss of order? Conversely, what is the critical inherent bending stiffness that a chain must possess in order to self-assemble into ordered macrostructures? The $\kappa-\epsilon$ phase diagram suggests (Figure 1) that while it is true that an intrinsic stiffness of the chain is a prerequisite for ordered aggregation, the property to self-assemble into these structures is not restricted to highly rigid polymers alone. Rather, even polymer chains for which $L_p < L$ can self-assemble into nematically ordered structures under certain parameter regimes (Figure 1 and 2). These findings suggest that the diversity in aggregated states and the rich phases observed in Figure 1 are a result of an interplay among three factors, viz, the strength of intermolecular interactions (ϵ), the propensity of the chain to assume extended structures (κ) and thermal fluctuations. These factors can be considered as experimental handles that could be tuned in order to drive polymer chains to assemble into various structures. Alteration to the microenvironment of the peptide could manifest itself as a change in the two phase-parameters (κ and ϵ) and as a result, the nature of the aggregated state. Our findings thus provide a general physical framework that is not only a basis for understanding the phenomenon of amyloidogenesis but also protein aggregation in general.

Our model also elucidates the role of the interaction parameter ϵ on the concentration dependence of aggregation. The ϵ-concentration phase-space shows that at higher values of ϵ, aggregation is favored even at lower concentrations (Supporting Information Figure S2 and S3). This suggests that the critical concentration at which self-assembly would occur can be lowered by strengthening the intermolecular interactions. Therefore, strengthening the interactions leads to an increased aggregation propensity, a behavior that is similar to the effect of concentration on the aggregation landscape in earlier studies.[78,79] It is thus evident from our simple model that the balance between aggregation and the entropic forces that favor disassembly could be tuned by increasing either the free monomer concentration or the strength of interaction.

Coexistence of Amorphous Aggregates and Ordered Structures: A Rationale for "Off-Pathway" Structures? The critical stiffness for which the transition from the disordered globular structures to the ordered assemblies would occur, is crucial in the context of amyloid aggregation. We observe that, for strongly interacting polymer chains (ϵ = 0.75 and 1 kcal/mol), the transition (with increasing κ) from the disordered aggregate state, to the ordered structures is abrupt and sharp (Figure 2B). This suggests that when the self-association tendencies are high enough, a slight increase in the propensity to access the extended configuration could result in a dramatic increase in order. In this regime of κ and ϵ, slight perturbations (mutations or change in microenvironment) could result in vastly different aggregated forms. The bimodal nature of the distribution near the transition point (Figure 2A, ϵ = 0.75 kcal/mol, black curve) further suggests that the systems at the edge of this disorder−order transition could assume both the ordered (fibril-like) and disordered (amorphous) states. This result signifies that slight variations in conditions in vitro and in vivo, in this intermediate regime, would result in the self-assembling system exhibiting high variability in both the kinetics and the end-state of the aggregate. This is an intriguing result in the light of amyloid aggregation experiments where amorphous aggregates are often considered to be one of the "off-pathway" structures that are secondary products of aggregation.[80,81] Our results suggest that under certain regimes of the phase parameters (ϵ and κ in this case), the experimental system would not necessarily conform to a single "pathway" that leads to ordered fibrillar entities alone. This is further supported by experimental studies that suggest the existence of multiple competing pathways of protein aggregation leading to formation of amyloids or other aggregation products.[55,80−83] For instance, a selective stabilization of one of the two unfolded intermediates of immunoglobulin has been reported to result in aggregates that are drastically different in order.[84] Also, small molecule amyloid inhibitors like reserveratrol have been used to stabilize the off-pathway oligomeric structures, which are reported to be slow-growing and thereby limiting amyloid growth.[85,86] Overall, using the idea of order−disorder transition for varying phase parameters, we provide an alternative explanation for the competing pathways in protein self-assembly that might result in different end-products of aggregation.

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

Polymer Stiffness Alone Does Not Ensure Aggregate Order. Stretches with β-sheet propensity are prevalent across the proteome.[21] However, the phenomenon of amyloid formation is not widespread in comparison. This could be attributed to various factors, primarily the aggregate clearance mechanisms and the sequestering of amyloidogenic sequences within the deep pockets of globular proteins.[21] However, some of the factors preventing amyloid formation could also lie in the design of peptide/protein sequences. Here, we probed the possible role of the length of the interacting patch in limiting ordered aggregation. By systematically varying the length of the polymer chains involved in self-assembly, we establish that stiffness of the chain alone is not a sufficient condition for ordered aggregation (Figure 4). Our results suggest that shorter chains do not assume ordered geometries despite their lengths being smaller than typical bending length-scales. Also, for a stiff homopolymer of a particular length, there exists a cricital interaction strength below which the aggregates remain disordered. For any value of ϵ, there is thus an optimal length at which order within the aggregate is maximum: short and very long protein chains ($L \gg L_p$) would collapse into disordered aggregates. Yanagi et al., in an aggregation study of human islet amyloid polypeptide (hIAPP) propose a similar length dependence of aggregation wherein an optimal length is favored for amyloid formation, with long and very short peptides disfavoring order.[87] These experimental findings lend further support to our claim that common signatures of protein aggregation could emerge from properties intrinsic to the polymer chain. For any peptide or a stretch of residues within a larger protein ("core sequence") to qualify as an amyloidogenic segment, it must possess the following properties; (i) sequence of residues with the ability to assume a linear geometry, (ii) optimal length of the sequence, and, (iii) a critical self-interaction strength for self-association. The absence of any of these three features would result in a significant reduction in the amyloidogenic propensity of the peptide chain. It is thereby possible that despite a large number of sequences being β-sheet compatible, amyloid aggregation is not rampant within the cell. We propose that the criticality in the length of the interaction patch could act as one of nature's regulatory mechanisms to prevent amyloid formation. This prediction is further supported by the length distribution of hydrophobic blocks within globular proteins, which suggests that long hydrophobic patches occur with very low frequencies.[88] This regulation in hydrophobic patch lengths could thus act as a deterrent against amyloid formation by proteins/peptides.

Extending the Model to Realistic Heteropolymer Aggregation Scenarios. Our homopolymer self-assembly simulations reveal that the aggregated state is the result of an interplay between polymer flexibility and interaction propensity. However, proteins/peptides are heteropolymers with unique sequences. Moreover, the sequence-dependence of amyloid formation has been established previously.[21,65] The sequence dependence of aggregation could arise due to multiple factors; the side-chain geometries of the amino acids and their compatibility within the β-sheet rich structure being the most crucial. We further extend our model to probe heteropolymeric aggregation that is typically observed in biology. We attempt to understand whether the sequence dependence partially emerges due to the arrangement of interactions within the aggregate structure, assuming that all the monomers are compatible with the extended configuration. Our simulations involving polymer chains with three bead types (P, N and H beads in Figure 5) clearly show that a mere shuffling of the sequence results in a loss of ability to assemble into ordered structures. A plausible reason for this sequence-dependence is that the self-assemblies are stabilized by short-range interactions. Any changes in the sequence, effected by the shuffling could result in a loss of the stable interaction network required to form linear aggregates. While some sequences lend an arrangement of interactions that can stabilize ordered macrostructures, the others with altered interaction networks are more stable in the globular form. Overall, our sequence study suggests that a clustering of the core self-interacting residues within the sequence is a key driving force for ordered self-assembly. Our results concur with the presence of an interaction core within diverse amyloidogenic peptides/proteins, which are known to drive the process of amyloidogenesis.[89−91]

Our study therefore provides a basic set of features that are essential for ordered aggregation to be favored. We employed these set of governing rules to develop a toy model for Aβ42, by defining regions of varying stiffness and self-interaction strength along the length of the polymer chain. These regions, defined on the basis of previous structural findings can reproduce the basic signatures of Aβ42 aggregation. Overall, our study provides the ground rules for ordered aggregation that could reproduce the basic features of amyloid aggregation and protein aggregation in general.

A Rationale for Designing Amyloidogenic Peptides. Amyloid fibrils possess an extremely robust structure, that is resistant to factors such as temperature, proteinase-K degradation and pressure in addition to possessing high tensile strength.[92−94] These properties of amyloids make them attractive bionanomaterials with potentially varied applications.[95−97] There has thus been great interest in designing peptides with enhanced self-assembly in order to exploit the favorable characteristics of amyloid fibrils. The current study provides guiding principles for polypeptide chains to self-assemble into highly ordered aggregates. This understanding could serve as a rationale for de novo synthesis of amyloidogenic peptides. The $\kappa-\epsilon$ phase-space highlights the interplay between the propensity of the peptide to access extended conformations and the strength of the intermolecular interactions in order to form ordered structures. A key aspect is to ensure that the sequences contain a combination of self-complementarity (stabilizing interactions, hydrophobic core) as well as the ability to access extended conformations with greater affinity. Sequences with an amalgamation of these two properties would possess a high amyloidogenic potency. However, as seen from our results, the length of the segment could act as both a promoting factor or a limiting factor for ordered aggregation. We propose that strongly self-interacting sequences of length 6 to 10 residues with β-sheet propensity would be optimal for ordered aggregation. While these factors alone may not be sufficient conditions for amyloid-like assembly, they would be prerequisites for such aggregation to be favored.

■ CONCLUSION

Self-association of proteins/peptides into supramolecular assemblies of varying structures has wide implications in biology. The diversity in sequence, structure and native functions of these self-assembling polypeptides suggests that the phenomenon is not restricted to any class of proteins/peptides. The hypothesis that amyloid formation could be a

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

generic property of the polypeptide chain is reinforced by studies where change in solution conditions could drive even typically nonaggregating proteins to self-assemble. In this study, we attempt to elucidate the signatures of amyloid formation that emerge out of the fundamental features of self-interacting polymers. By varying parameters like interaction strength, bending stiffness and polymer chain length, we capture various signatures like aggregate diversity, polymorphism, sequence dependence, etc., that are typically associated with the process of protein aggregation and amyloid formation. The fact that a minimal model of simple self-interacting polymers (with no explicit side-chain information) could give rise to these signatures is a strong vindication for amyloid formation being a generic property of all polypeptide chains. While the inclusion of details like side-chain information would give rise to some finer features of the phenomenon and capture structural details with greater accuracy, our simple model nevertheless allows us to identify the key factors governing aggregation. This fundamental understanding could be vital not only as a basis for designing control strategies against amyloid-associated diseases but also for manipulating proteins/peptides as self-assembling entities.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b06433.

Alternative representations of the self-assembled structures; results of concentration-dependence simulations; the length-ϵ phase diagram; the results of STLNFE peptide self-assembly simulation in comparison with electron micrographs (PDF)
Video S1 (MPG)
Video S2 (MPG)
Video S3 (MPG)
Video S1−S3 captions (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*srivastav_r@iitb.ac.in
*samirmaji@iitb.ac.in
*ranjithp@iitb.ac.in

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Prof. Dibyendu Das, Prof. Sunil Kumar and Prof. Shamik Sen for their suggestions during the conceptualization of the work. They would also like to thank Aparna JS and Anoop Arunagiri for their input while writing the paper. The authors are grateful to Prof. Luis Serrano for allowing them to use their electron micrographs in the Supporting Information to support their findings.

■ REFERENCES

(1) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.
(2) Maji, S. K.; Perrin, M. H.; Sawaya, M. R.; Jessberger, S.; Vadodaria, K.; Rissman, R. a.; Singru, P. S.; Nilsson, K. P. R.; Simon, R.; Schubert, D.; Eisenberg, D.; Rivier, J.; Sawchenko, P.; Vale, W.; Riek, R. *Science* **2009**, *325*, 328−32.
(3) Dogterom, M.; Yurke, B. *Science* **1997**, *278*, 856−860.
(4) Dogterom, M.; Kerssemakers, J. W. J.; Romet-Lemonne, G.; Janson, M. E. *Curr. Opin. Cell Biol.* **2005**, *17*, 67−74.
(5) Footer, M. J.; Kerssemakers, J. W. J.; Theriot, J. A.; Dogterom, M. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2181−2186.
(6) Inoué, S.; Salmon, E. D. *Mol. Biol. Cell* **1995**, *6*, 1619−1640.
(7) Mogilner, A.; Oster, G. *Biophys. J.* **1996**, *71*, 3030−3045.
(8) Fändrich, M. *J. Mol. Biol.* **2012**, *421*, 427−440.
(9) Bucciantini, M.; Calloni, G.; Chiti, F.; Formigli, L.; Nosi, D.; Dobson, C. M.; Stefani, M. *J. Biol. Chem.* **2004**, *279*, 31374−31382.
(10) Greenwald, J.; Riek, R. *Structure* **2010**, *18*, 1244−60.
(11) Tanaka, M.; Komi, Y. *Nat. Chem. Biol.* **2015**, *11*, 373−377.
(12) Pellarin, R.; Guarnera, E.; Caflisch, A. *J. Mol. Biol.* **2007**, *374*, 917−924.
(13) Qiao, Q.; Bowman, G. R.; Huang, X. *J. Am. Chem. Soc.* **2013**, *135*, 16092−16101.
(14) Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L. M.; Dobson, C. M.; Vendruscolo, M.; Favrin, G. *PLoS Comput. Biol.* **2007**, *3*, e173.
(15) Neupane, K.; Solanki, A.; Sosova, I.; Belov, M.; Woodside, M. T. *PLoS One* **2014**, *9*, e86495.
(16) Jha, N. N.; Anoop, A.; Ranganathan, S.; Mohite, G. M.; Padinhateeri, R.; Maji, S. K. *Biochemistry* **2013**, *52*, 8800−10.
(17) Anoop, A.; et al. *J. Biol. Chem.* **2014**, *289*, 16884−16903.
(18) Chiti, F.; Webster, P.; Taddei, N.; Clark, a.; Stefani, M.; Ramponi, G.; Dobson, C. M. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 3590−3594.
(19) Lashuel, H. A.; Lansbury, P. T. *Q. Rev. Biophys.* **2006**, *39*, 167−201.
(20) Dobson, C. M. *Nature* **2003**, *426*, 884−890.
(21) Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 3487−3492.
(22) Corrigan, A. M.; Muller, C.; Krebs, M. R. H. *J. Am. Chem. Soc.* **2006**, *128*, 14740−14741. PMID: 17105248.
(23) Cannon, D.; Donald, A. M. *Soft Matter* **2013**, *9*, 2852−2857.
(24) Pawar, A. P.; DuBay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. *J. Mol. Biol.* **2005**, *350*, 379−392.
(25) Tjernberg, L.; Hosia, W.; Bark, N.; Thyberg, J.; Johansson, J. *J. Biol. Chem.* **2002**, *277*, 43243−43246.
(26) Bellesia, G.; Shea, J.-E. *J. Chem. Phys.* **2009**, *130*, 145103.
(27) Chen, S.; Gao, S.; Cheng, D.; Huang, J. *Biochem. Biophys. Res. Commun.* **2014**, *447*, 255−262.
(28) Ma, B.; Nussinov, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14126−14131.
(29) Ma, B.; Nussinov, R. *Protein Sci.* **2002**, *11*, 2335−2350.
(30) Klimov, D. K.; Thirumalai, D. *Structure* **2003**, *11*, 295−307.
(31) Rohrig, U. F.; Laio, A.; Tantalo, N.; Parrinello, M.; Petronzio, R. *Biophys. J.* **2006**, *91*, 3217−3229.
(32) Buchete, N.-V.; Tycko, R.; Hummer, G. *J. Mol. Biol.* **2005**, *353*, 804−821.
(33) Buchete, N.-V. *Biophys. J.* **2012**, *103*, 1411−1413.
(34) Vacha, R.; Frenkel, D. *Biophys. J.* **2011**, *101*, 1432−1439.
(35) Zierenberg, J.; Janke, W. *EPL* **2015**, *109*, 28002.
(36) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1−19.
(37) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819−834. PMID: 26621095.
(38) Cuesta, I. G.; Sanchez de Meras, A. M. J. *Phys. Chem. Chem. Phys.* **2014**, *16*, 4369−4377.
(39) Narang, P.; Bhushan, K.; Bose, S.; Jayaram, B. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2364−2375.
(40) Danielsson, J.; Andersson, A.; Jarvet, J.; Gräslund, A. *Magn. Reson. Chem.* **2006**, *44*, S114−S121.
(41) Chin, A.; Toptygin, D.; Elam, W.; Schrank, T.; Hilser, V. *Biophys. J.* **2016**, *110*, 362−371.
(42) Sanchez-Castillo, A.; Osipov, M. A.; Giesselmann, F. *Phys. Rev. E* **2010**, *81*, 021707.
(43) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. *Bioinformatics* **2007**, *23*, 2625−2627.
(44) Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. *J. Chem. Phys.* **2004**, *121*, 10748.

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922

(45) Horbach, J.; Kob, W. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *60*, 3169−3181.
(46) Levine, B. G.; Stone, J. E.; Kohlmeyer, A. *J. Comput. Phys.* **2011**, *230*, 3556−3569.
(47) Moe, N. E.; Ediger, M. D. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1999**, *59*, 623−630.
(48) Di Michele, L.; Fiocco, D.; Varrato, F.; Sastry, S.; Eiser, E.; Foffi, G. *Soft Matter* **2014**, *10*, 3633−3648.
(49) Hansen, J.-P.; McDonald, I. R. In *Theory of Simple Liquids*, 3rd ed.; Hansen, J.-P., McDonald, I. R., Eds.; Academic Press: Burlington, 2006; pp 78−108.
(50) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(51) Nguyen, H. D.; Hall, C. K. *Biophys. J.* **2004**, *87*, 4122−4134.
(52) Nguyen, H. D.; Hall, C. K.; Carolina, N. *J. Biol. Chem.* **2005**, *280*, 9074−9082.
(53) Zanuy, D.; Ma, B.; Nussinov, R. *Biophys. J.* **2003**, *84*, 1884−1894.
(54) Baskakov, I. V.; Legname, G.; Baldwin, M. A.; Prusiner, S. B.; Cohen, F. E. *J. Biol. Chem.* **2002**, *277*, 21140.
(55) Huang, T. H. J.; Yang, D.-S.; Fraser, P. E.; Chakrabartty, A. *J. Biol. Chem.* **2000**, *275*, 36436−36440.
(56) Goldsbury, C. S.; Cooper, G. J.; Goldie, K. N.; Muller, S. A.; Saafi, E. L.; Gruijters, W.; Misur, M. P.; Engel, A.; Aebi, U.; Kistler, J. *J. Struct. Biol.* **1997**, *119*, 17−27.
(57) Meinhardt, J.; Sachse, C.; Hortschansky, P.; Grigorieff, N.; Fandrich, M. *J. Mol. Biol.* **2009**, *386*, 869−877.
(58) Onsager, L. *Ann. N. Y. Acad. Sci.* **1949**, *51*, 627−659.
(59) Gupta, A. M.; Edwards, S. F. *J. Chem. Phys.* **1993**, *98*, 1588.
(60) Ghosh, K.; Carri, G. A.; Muthukumar, M. *J. Chem. Phys.* **2002**, *116*, 5299.
(61) Doi, M. *Soft Matter Physics*; Oxford University Press: Oxford, 2013; p 272.
(62) Das, R. K.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13392−13397.
(63) Sawle, L.; Ghosh, K. *J. Chem. Phys.* **2015**, *143*, 085101.
(64) Bergen, M. V.; Friedhoff, P.; Biernat, J.; Heberle, J.; Mandelkow, E.; Mandelkow, E. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5129−5134.
(65) Lopez de la Paz, M.; Serrano, L. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 87−92.
(66) Lakshmanan, A.; Cheong, D. W.; Accardo, A.; Di Fabrizio, E.; Riekel, C.; Hauser, C. A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 519−524.
(67) Yoshimura, Y.; Lin, Y.; Yagi, H.; Lee, Y.-H.; Kitayama, H.; Sakurai, K.; So, M.; Ogi, H.; Naiki, H.; Goto, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 14446−14451.
(68) Garai, K.; Sengupta, P.; Sahoo, B.; Maiti, S. *Biochem. Biophys. Res. Commun.* **2006**, *345*, 210−215.
(69) He, J.; Xing, Y.-F.; Huang, B.; Zhang, Y.-Z.; Zeng, C.-M. *J. Agric. Food Chem.* **2009**, *57*, 11391−11396. PMID: 19904937.
(70) Shammas, S.; Knowles, T.; Baldwin, A.; MacPhee, C.; Welland, M.; Dobson, C.; Devlin, G. *Biophys. J.* **2016**, *100*, 2783−2791.
(71) Luhrs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Dobeli, H.; Schubert, D.; Riek, R. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17342−17347.
(72) Fernandez-Escamilla, A.-M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. *Nat. Biotechnol.* **2004**, *22*, 1302−1306.
(73) Laradji, M.; Sunil Kumar, P. B. *J. Chem. Phys.* **2005**, *123*, 224902.
(74) Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2011**, *133*, 14160−14163.
(75) Morel, B.; Varela, L.; Azuaga, A. I.; Conejero-Lara, F. *Biophys. J.* **2010**, *99*, 3801−3810.
(76) Gosal, W. S.; Morten, I. J.; Hewitt, E. W.; Smith, D. A.; Thomson, N. H.; Radford, S. E. *J. Mol. Biol.* **2005**, *351*, 850−864.
(77) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. *Nature* **2002**, *416*, 507−511.
(78) Schmit, J. D.; Ghosh, K.; Dill, K. *Biophys. J.* **2011**, *100*, 450−458.
(79) Lee, C. F. *Phys. Rev. E* **2009**, *80*, 031922.
(80) Hall, D.; Kardos, J.; Edskes, H.; Carver, J. A.; Goto, Y. *FEBS Lett.* **2015**, *589*, 672−679.
(81) Powers, E. T.; Powers, D. L. *Biophys. J.* **2008**, *94*, 379−391.
(82) Gorman, P. M.; Yip, C. M.; Fraser, P. E.; Chakrabartty, A. *J. Mol. Biol.* **2003**, *325*, 743−757.
(83) Vetri, V.; Canale, C.; Relini, A.; Librizzi, F.; Militello, V.; Gliozzi, A.; Leone, M. *Biophys. Chem.* **2007**, *125*, 184−190.
(84) Qin, Z.; Hu, D.; Zhu, M.; Fink, A. L. *Biochemistry* **2007**, *46*, 3521−3531. PMID: 17315948.
(85) Nedumpully-Govindan, P.; Kakinen, A.; Pilkington, E. H.; Davis, T. P.; Chun Ke, P.; Ding, F. *Sci. Rep.* **2016**, *6*, 19463.
(86) Ladiwala, A. R. A.; Lin, J. C.; Bale, S. S.; Marcelino-Cruz, A. M.; Bhattacharya, M.; Dordick, J. S.; Tessier, P. M. *J. Biol. Chem.* **2010**, *285*, 24228−24237.
(87) Yanagi, K.; Ashizaki, M.; Yagi, H.; Sakurai, K.; Lee, Y.-H.; Goto, Y. *J. Biol. Chem.* **2011**, *286*, 23959−23966.
(88) Schwartz, R.; Istrail, S.; King, J. *Protein Sci.* **2001**, *10*, 1023−1031.
(89) Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. *Biochemistry* **2000**, *39*, 13748−13759. PMID: 11076514.
(90) Tenidis, K.; Waldner, M.; Bernhagen, J.; Fischle, W.; Bergmann, M.; Weber, M.; Merkle, M.-L.; Voelter, W.; Brunner, H.; Kapurniotu, A. *J. Mol. Biol.* **2000**, *295*, 1055−1071.
(91) Reches, M.; Porat, Y.; Gazit, E. *J. Biol. Chem.* **2002**, *277*, 35475−35480.
(92) Smith, J. F.; Knowles, T. P. J.; Dobson, C. M.; MacPhee, C. E.; Welland, M. E. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15806−15811.
(93) Meersman, F.; Dobson, C. M. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 452−460. Proteins Under High Pressure.
(94) Mesquida, P.; Riener, C. K.; MacPhee, C. E.; McKendry, R. A. *J. Mater. Sci.: Mater. Med.* **2007**, *18*, 1325−1331.
(95) Gras, S. L. In *Engineering Aspects of Self-Organizing Materials*; Koopmans, R. J., Ed.; Academic Press, 2009; Advances in Chemical Engineering, Vol. 35, pp 161−209.
(96) Pertinhez, T. A.; Conti, S.; Ferrari, E.; Magliani, W.; Spisni, A.; Polonelli, L. *Mol. Pharmaceutics* **2009**, *6*, 1036−1039.
(97) Hamedi, M.; Herland, A.; Karlsson, R. H.; Inganäs, O. *Nano Lett.* **2008**, *8*, 1736−1740.

DOI: 10.1021/jacs.6b06433
J. Am. Chem. Soc. 2016, 138, 13911−13922