Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly

Wei-Feng Xue, Steve W. Homans, and Sheena E. Radford*

Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

Edited by William A. Eaton, National Institutes of Health, Bethesda, MD, and approved April 7, 2008 (received for review December 11, 2007)

Self-assembly of misfolded proteins into ordered fibrillar aggregates known as amyloid results in numerous human diseases. Despite an increasing number of proteins and peptide fragments being recognised as amyloidogenic, how these amyloid aggregates assemble remains unclear. In particular, the identity of the nucleating species, an ephemeral entity that defines the rate of fibril formation, remains a key outstanding question. Here, we propose a new strategy for analyzing the self-assembly of amyloid fibrils involving global analysis of a large number of reaction progress curves and the subsequent systematic testing and ranking of a large number of possible assembly mechanisms. Using this approach, we have characterized the mechanism of the nucleation-dependent formation of β_2-microglobulin (β_2m) amyloid fibrils. We show, by defining nucleation in the context of both structural and thermodynamic aspects, that a model involving a structural nucleus size approximately the size of a hexamer is consistent with the relatively small concentration dependence of the rate of fibril formation, contrary to expectations based on simpler theories of nucleated assembly. We also demonstrate that fibril fragmentation is the dominant secondary process that produces higher apparent cooperatively in fibril formation than predicted by nucleated assembly theories alone. The model developed is able to explain and predict the behavior of β_2m fibril formation and provides a rationale for explaining generic properties observed in other amyloid systems, such as fibril growth acceleration and pathway shifts under agitation.

AIC model comparison analysis | amyloid fibril formation | fibril brittleness | global analysis

Self-assembly of misfolded forms of normally soluble and functional proteins or peptides into amyloid fibrils results in numerous human diseases (1). Understanding how amyloid self-assembly occurs, therefore, is of paramount importance for a molecular interpretation of amyloidosis and for the development of therapies against amyloid disease. Over the past decade, advances have been made toward a more complete description of amyoid fibril formation, including the determination of increasingly refined models of fibril structures (reviewed in ref. 1) and the identification of amyloid precursors and oligomeric states reviewed in ref. 2, one or more of which could be the culprits of cytotoxicity associated with several amyloid diseases (e.g., ref. 3). However, the molecular events occurring during the self-assembly process itself remain obscure because of the heterogeneity and the complexity of the early association events.

Amyloid fibril self-assembly reactions are generally accepted as a form of nucleated polymerization (4). These reactions are characterized by an initial lag phase where little or no change in fibril concentration can be detected. This is followed by an elongation phase where a large mass percentage of the starting protein material is converted into fibrils. A shared feature among amyloid formation and other nucleated processes is that the lag phase can be partly or entirely bypassed by the addition of seeds. In the case of amyloid fibrils, seeds are usually fragments of preformed fibrils. Traditionally, information regarding the species present during the lag phase, for example, the size of the so-called "critical nucleus," are gained through analysis of the protein concentration dependence of the rate of fibril formation, determined for example by the time to reach 50% reaction completion (t_{50}), the length of the lag phase (t_{lag}), or the initial rate of fibril formation (4). This simple approach is based on nucleation-dependent polymerization theories and other similar and related models originally developed for actin and sickle cell haemoglobin assembly (4, 5) and dictates an increasingly strong concentration dependence of these parameters with an increase in the size of the "critical nucleus." Many studies have applied this approach to amyloid assembly reactions, with the majority of cases showing a very small concentration dependence of the reaction progress (e.g., refs. 6–11), leading to the conclusion that the "critical nucleus" is monomeric or very small. Although simple to perform, mounting evidence indicates that this type of analysis is inadequate because of the presence of large populations of prenucleus oligomers (12, 13) and/or the use of mechanical agitation during fibril growth (8, 14), which can diminish the apparent concentration dependence of fibril growth leading to an underestimation of the nucleus size(s). Another drawback of such an analysis is that the sigmoidal shape of the progress curves are usually ignored, even though fibril formation most often displays a higher degree of apparent cooperativity than is typically theorised and observed in other nucleated reactions [supporting information (SI) Fig. S1], and upon which the amyloid assembly models were drawn (4).

Here, we provide a new strategy for the analysis of amyloid fibril formation by acquisition of >200 fibrillation progress curves under the widest possible range of protein concentration by defining nucleation in the context of both structural and thermodynamic aspects and by developing a generalized modular modeling approach that allows a variety of possible self-assembly mechanisms to be tested systematically. Using this approach, we have studied the assembly of β_2-microglobulin (β_2m) into amyloid-like fibrils at low pH (15, 16). Under these conditions, fibril formation occurs both rapidly and quantitatively, without visible formation of amorphous aggregates, providing an ideal system with which to develop theories of nucleated amyloid assembly.

Results

Experimental Measurement of the Concentration Dependence of Fibril Formation. To obtain information regarding the nature of the nucleating species in β_2m amyloid formation, a large dataset

Author contributions: W.-F.X., S.W.H., and S.E.R. designed research; W.-F.X. performed research; W.-F.X. analyzed data; and W.-F.X., S.W.H., and S.E.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

*To whom correspondence should be addressed. E-mail: s.e.radford@leeds.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0711664105/DCSupplemental.

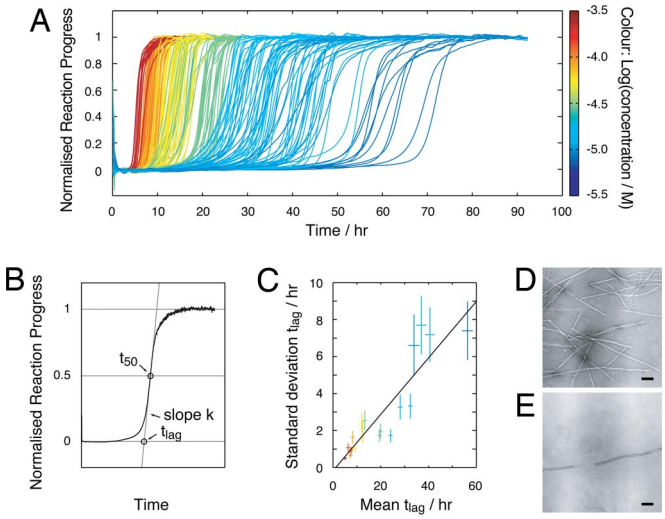

Fig. 1. Reaction progress curves of amyloid fibril formation of β_2m. (*A*) Reaction progress curves monitored by ThT fluorescence. A total of 235 traces are shown at 20 different protein concentrations ranging from 8 to 244 μM (color bar on the *Right*). (*B*) Representative illustration of the method by which t_{lag} and k are extracted from the reaction traces. (*C*) The variation in the replicates shown as the standard deviation of t_{lag} plotted vs. the average t_{lag}. The error bars represent the standard error of the standard deviation and the standard error of the mean, respectively. A free linear fit is also shown. (*D* and *E*) Negative stain EM images of samples after the reaction reached completion obtained at a protein concentration of 244 μM (*D*) or 8 μM (*E*). (Scale bars, 100 nm.)

consisting of 235 progress curves was obtained at monomer concentrations ranging from 8 to 244 μM monitored by the fluorescence of the amyloid-specific dye thioflavin T (ThT) (Fig. 1*A*). Twelve replicate reaction traces were typically collected at each protein concentration. In all cases, the samples were agitated (see *Materials and Methods*) to facilitate rapid fibril growth. Each curve displays lag-phase kinetics typical of nucleated growth. ThT fluorescence was chosen as the reporter to monitor the reaction progress because of its high sensitivity and signal to noise ratio (Fig. S2), which became particularly important at low protein concentrations. Control experiments, using intrinsic tryptophan fluorescence, light scattering, turbidity and EM, showed that the binding of ThT does not perturb the rate of assembly and confirmed that ThT reports on the same fibril formation process as the other detection methods (Fig. S2). Negative stain EM showed, in addition, that the fibrils formed under every condition used have a long-straight morphology typical of amyloid (Fig. 1 *D* and *E*). Principal parameters describing the transitional regions of the progress curves: The lag time (t_{lag}) and the transition slope (k) were extracted from the normalised progress curves (see *Materials and Methods* and Fig. 1*B*) for subsequent analysis.

As shown in Fig. 1*A* (as expected for a nucleated reaction), the length of the lag-time is increased as the protein concentration is decreased. However, at each protein concentration, a significant variation in the lag time is observed. This is particularly obvious at the lowest concentrations used (dark blue traces in Fig. 1*A*). The standard deviation of t_{lag} at each protein concentration is plotted against the average of t_{lag} in Fig. 1*C*. The plot shows that the variation in t_{lag} is linearly correlated with the average, with the fitted intercept close to zero, demonstrating the random nature of the variation, independent of the average. This suggests that the variations observed in the data are likely to originate from subtle experimental differences between the replicates, but may also reflect contributions from the stochastic nature of the nucleation mechanism (17, 18). Another important feature of the data is that all of the curves have steep transitions similar to the progress curves observed for the assembly of fibrils from other amyloidogenic proteins and peptides (e.g., refs. 8 and 11) but distinctively different from those predicted by nucleated assembly models alone (4, 5) (Fig. S1). The steepness of the transitions indicates the presence of a secondary process accelerating the reaction under the conditions used (4).

Construction of a Generic Mechanistic Modeling Approach for Fibril Self-Assembly. To gain new insights into the mechanism of amyloid formation, a model capable of describing the self-assembly process, including the progress curves and their concentration dependence, is required. Many different mechanistic models have been used to describe the assembly of amyloid and other self-assembled protein polymers (e.g., refs. 4, 5, 12, 14, and 19–23). Because of the complexity of the system and the consequent large number of different possibilities involved, a modular approach was used here to model amyloid assembly. We have rationalized the mechanism into three distinct modules: prepolymerization, polymerization, and secondary processes (Fig. 2). Each possible assembly model is then composed of a variant of each of the three modules (a full mathematical and thermodynamical description is given in *SI Text* and Table S1). Key features of the model design are: (*i*) The modular design enables a large number of different assembly mechanisms to be tested and ranked. (*ii*) The generic assembly reaction scheme used (Fig. 2*A* and *SI Text*, Eq. **1**) does not itself constrain the assembly model but allows the assembly mechanism to be derived by the choice of functions describing the rate constants and reaction free energies for each step of the assembly (here step, linear, and power functions were tested, each corresponding to a distinct assembly scenario where different free energy contributions, such as specific interactions or translation/rotational entropies, dominate; see Fig. 2*C* and *SI Text*). (*iii*) The structural and thermodynamic aspects of assembly are separately and exactly defined, with the shape of the reaction free energies for each assembly step as function of the extent of assembly defining the structural nucleus size n_S (*SI Text*, Eq. **9**). The free energy for each assembly species compared with the monomer (taking into account the contribution of translational entropy loss by reduction in the number of particles during assembly) then defines the thermodynamic nucleus size n_T (*SI Text*, Eq. **10**). (*iv*) A secondary fragmentation process that includes a length and position-dependent rate of fragmentation is used (Fig. 2*D*, *SI Text*, and Eq. **13**). (*v*) The progress curves for different possible mechanisms are calculated numerically, which eliminates the need for additional, potentially erroneous, assumptions used in analytical models such as negligible oligomer populations and irreversible steps (4, 12). In summary, the key idea behind our approach is that, instead of using a single model containing simplifications that are based on predetermined assumptions that may be difficult to justify or validate in the analysis of experimental data, we use a flexible approach that allows a large number of possible models to be tested against the experimental data.

Comparing Model Variants. Next, we applied our modeling approach to the analysis of the experimental fibril growth curves for β_2m depicted in Fig. 1. The transitional regions described by the curve parameters t_{lag} and k were extracted from the numerical solutions of the models in the same way as for the experimental reaction progress traces shown in Fig. 1*B*. The resulting values were then compared with the experimental data through least-square analysis in order that the model(s) best representing the experimental progress curves can be determined. Key to this analysis were: (*i*) the acquisition of data over the largest possible protein concentration range (limited by the sensitivity of the assay and the need to avoid nonspecific aggregation); (*ii*) the

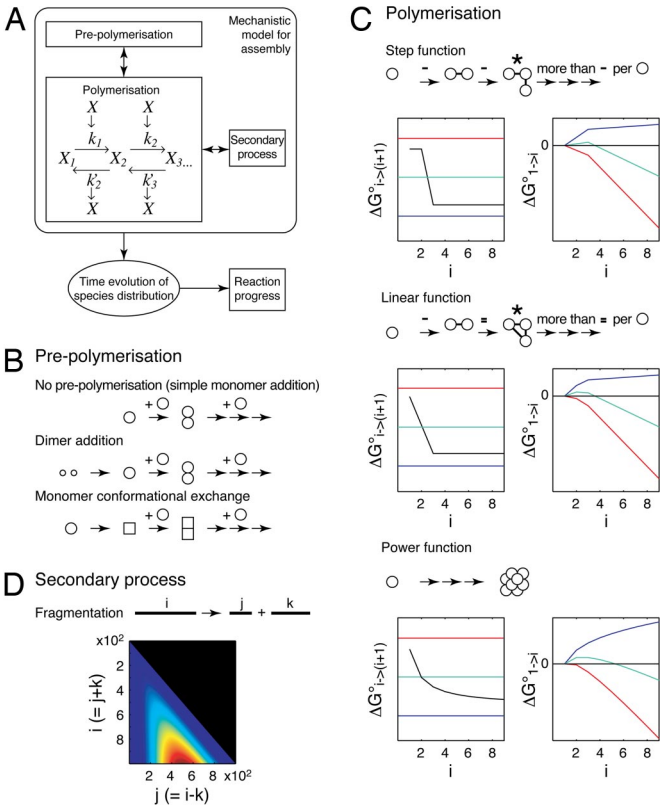

Fig. 2. Schematic illustration of the modeling strategy. (*A*) The overall model configuration consists of three modules: prepolymerization, polymerization, and secondary processes. The primary output is the distribution of species as a function of time. This distribution is then used to calculate reaction progress curves and other quantities that can be used to compare with the experimental data. (*B*) Diagram representation of the three different prepolymerization modules tested. (*C*) Different reaction free energy functions tested. For each case, the reaction free energy of each addition step as function of the extent of assembly are shown on the *Left*, and the free energy difference compared with the monomer is shown on the *Right*. Colored lines illustrate the reaction free energy $\Delta G^{\circ}_{i \rightarrow (i+1)} = 0$ when the loss in translational entropy due to the reduction of particles at different concentrations is taken into account (*Left*), and the free energy of species compared with monomer (*Right*) at high (red), intermediate (green), or low (blue) initial monomer concentrations. A diagram describing the structural aspects of each variant tested is shown above each plot (−, interaction; *, structural nucleus in the case of step and linear functions). In the case of the power free energy function, which does not have a specific n_S, the intercept of the tangent to the initial slope and the value at $i \rightarrow \infty$ are used to estimated n_S. (*D*) Illustration of the model used to describe the secondary process of fragmentation. The contour plot shows the dependence of the fragmentation rate constant on the total fibril length *i* and the breakage point in the fibril *j* for fibril lengths up to 1,000. The relative magnitudes of the fragmentation rate constant are shown from low (blue) to high (red).

collection of sufficient data to enable accurate statistical analysis; and (*iii*) the confirmation that fibrils with identical morphology were obtained under all conditions. Most importantly, global analysis of the entire dataset was carried out with the numerically calculated progress curves rather than analysis with any single kinetic parameter. A total of 21 different models, constructed through different combinations of variants of the three modules described above, were tested (numbered in Table S2). The pairwise relationship plots of the t_{lag}, k, and initial monomer concentration, together with representative examples of globally fitted models are shown in Fig. 3, with each 2D plot showing different representations of the dataset and the model. Although such a representations provide a useful means of displaying the data and the predictions of different models, the goodness of fit of the different models was quantified and ranked through the calculation of Akaike information criterion (AIC) scores (see *SI Text*) that take into account the quality of the global fit and the number of fitted parameters (Table S2).

Through ranking of AIC scores of the 21 models tested, a single best-fit model was identified. This model involves monomer addition through a hexameric structural nucleus involving fragmentation as a secondary process (Fig. 3, column three, $n_S = 6$ and Table S2, model 5). Although significantly poorer in describing the experimental data, the models with the nearest best fits included a rarely populated monomeric species as the structural and thermodynamic nucleus (Fig. 3 leftmost column Table S2, model 21), and some models with n_S larger than one but with the n_T approaching one (Table S2, models 8–10). However, models without fragmentation as a secondary process are unable to describe the experimental data (rightmost column in Fig. 3 and Table S2, models 1, 12, and 20). Importantly, although a single model of the 21 tested emerged as best able to describe the data if the global picture is considered, analyses based solely on the concentration dependence of a single kinetic parameter, such as t_{lag} or t_{50}, are inadequate in distinguishing between different possible assembly mechanisms (Fig. 3 *Top*). This is especially true for amyloid assembly reactions displaying a high apparent cooperativity of assembly, which is indicative of the involvement of secondary processes, such as fragmentation (24).

To further distinguish the best-fit model from the other alternative possibilities, we next analyzed the size of species populated during the lag phase predicted by each model. The best-fit model involving step free energy function with $n_S = 6$ (Table S2, model 5) predicts a significant decrease in monomer concentration and a corresponding increase in the population of on-pathway oligomeric species preceding the hexameric structural nucleus in the lag phase (Fig. 4). Other models tested that have relatively good agreement with the data, such as the model involving a rarely populated monomeric species as the structural nucleus (Table S2, model 21), or $n_S = 4$ with linear free energy function (Table S2, model 8), predict monomer depletion that reflects the formation of fibrils and the absence of significant amounts of oligomers in the lag phase. Importantly, previous analysis of fibril growth under conditions very similar to those used here using ESI-MS demonstrated the presence of dimers, trimers, and tetramers in the lag phase, but no higher order species were found (25). The ESI-MS analysis also demonstrated a significant decrease in the monomer concentration in the lag phase, consistent with the predictions of the best-fit model (Table S2, model 5). Parallel experiments using sedimentation velocity AUC also revealed an equilibrium between 1 and 4 mers in the lag-phase, although rapid equilibrium between these species on the timescale of the experiments precluded determination of their relative populations (25). These experiments provide powerful independent information that mirrors the predictions made assuming the best-fit model with $n_S = 6$ (Fig. 4), but are inconsistent with the other models tested that fitted the fibril formation data less well. The best model able to describe all of the experimental data available, therefore, is that involving monomer addition to a hexameric structural nucleus in a step nucleating polymerization reaction with fragmentation as a secondary process.

Testing the Model Predictions. As a further test of the robustness of the above model, we next performed additional experiments and used the resulting information to test the predicted properties of the described system. First, we carried out fibril growth experiments in the absence of agitation but in the presence of fibril seeds. Fibrils used as seeds were formed at an initial monomer concentration of 63 μM exactly as described above. At the end of fibril growth (50 h), 2% (vol/vol) of the sample was

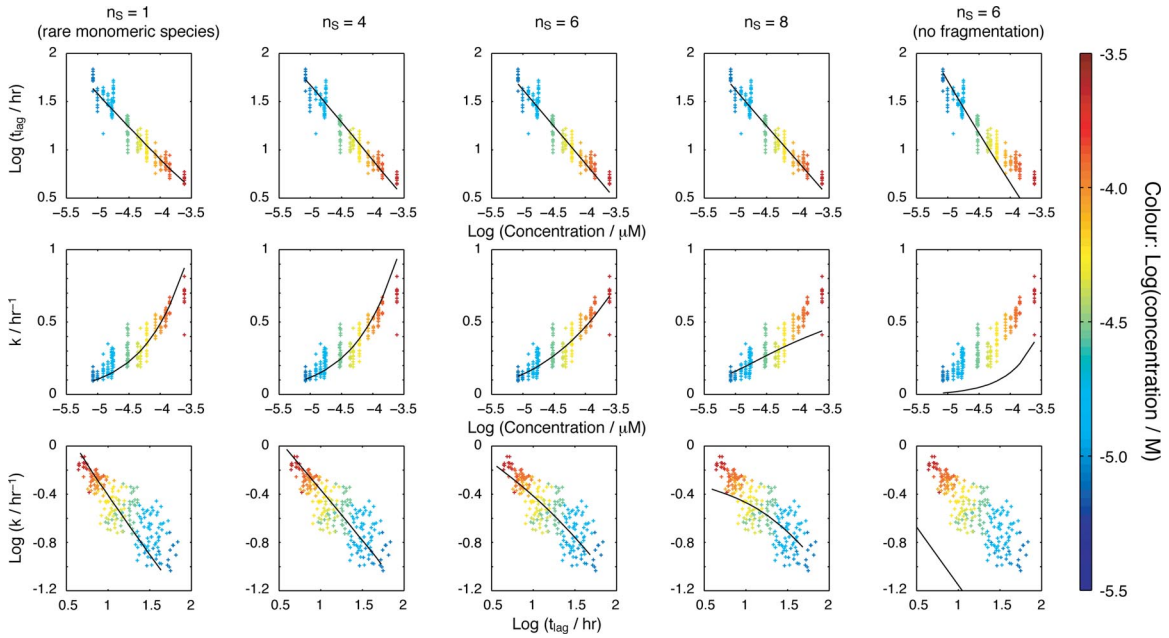

Fig. 3. The curve parameters t_{lag} and k representing the experimental data at different initial monomer concentrations shown in Fig. 1*A* (+) plotted as different pairwise combinations, together with selected global least-square fitted models shown in the same pairwise representations (black lines). (*Top*) log(concentration) vs. log(t_{lag}). (*Middle*) log(concentration) vs. k. (*Bottom*) log(t_{lag}) vs. log (k). All models shown involve monomer addition and a step nucleation polymerization mechanism and, with the exception of the rightmost plots, included fragmentation as a secondary process. Although each figure serves to represent the data and the predictions of different models, the best fit model was selected by global analysis and subsequent ranking of the goodness of each fit.

added directly to freshly prepared protein solutions to seed growth. The rate of fibril growth was monitored at different initial monomer concentrations ranging from 1.5 μM to 12 μM. The resulting reaction progress curves showed traces without a detectable lag phase and the reaction reached completion within 5–10 h (Fig. S3*A*). Importantly, the model that best describes the experimental data for the unseeded agitated growth reactions shown in Fig. 1*A* (monomer addition, step nucleation at $n_S = 6$, with fragmentation as a secondary process) reproduces accurately the progress curves obtained in the absence of agitation through seeded fibril growth (Fig. S3*B*). However, if fragmentation is excluded during seed formation, the model is no longer able to describe the experimental data (Fig. S3*B Inset*), instead predicting that even under seeded conditions fibril formation would take >500 days to complete.

Second, if fragmentation is the dominating secondary process during agitated fibril growth, as the best-fit model suggests, the ability of fibrils to seed new reactions should depend on the number of fibril ends, determined by the extent of fragmentation. By contrast, if the secondary process involves heterogeneous nucleation on fibril surface (26), then the seeding efficiency should be independent of the number of fibril ends, because the total surface area of thin linear aggregates will be essentially independent of agitation. To determine whether this is the case, fibrils were formed in the absence of agitation but in the presence of the same amount seeds by weight that had been fragmented to different extents by vigorous agitation (see *Materials and Methods*). The seeding efficiency characterized by the initial slope of the seeded reaction progress curves was then monitored (Fig. S3*C*). The data show that the seeding efficiency

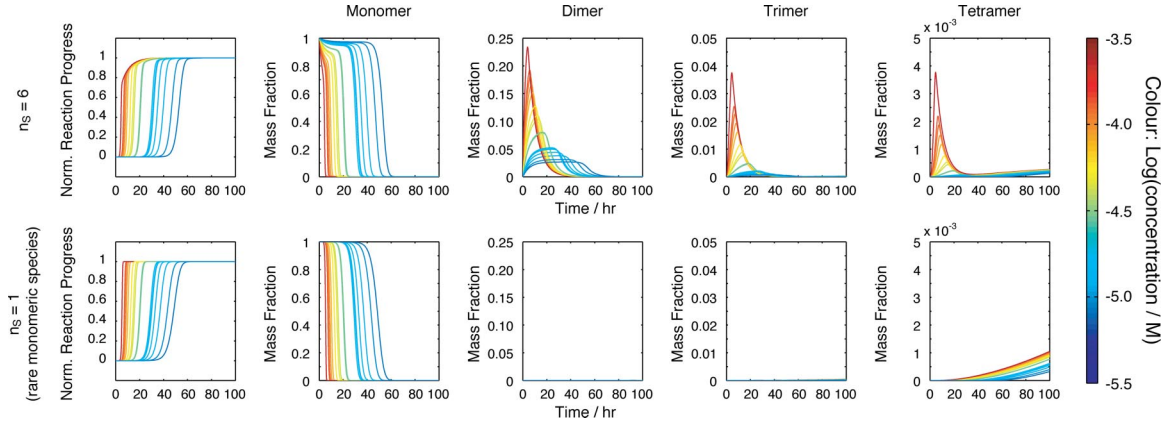

Fig. 4. Comparison of the calculated reaction progress (leftmost column), together with the population of monomer dimer, trimer and tetramer vs. time. (*Upper*) The best-fit model (monomer addition, step free energy function with $n_S = 6$, including secondary fragmentation) (Table S2, model 5). (*Lower*) Next-best model (monomer addition, rare monomeric structural nucleus and secondary fragmentation) (Table, model 21).

increases as the extent of fragmentation is increased, precisely as predicted assuming that fragmentation is the dominant secondary process under nonquiescent conditions.

Discussion

Polymerization Free Energy Profile and Structural vs. Thermodynamic Nucleus. In this work, we have developed a systematic approach for the analysis of the mechanism of nucleation-dependent assembly into amyloid fibrils. The approach described is generically applicable and enables the complexities of amyloid assembly mechanisms to be dissected. It also allows different oligomeric species to be identified and their population quantified, providing a structural and thermodynamic platform for the analysis of nucleated mechanisms and allowing rare oligomeric species to be predicted and their populations quantified. The approach involves comparing variants of a generic and modular mechanistic assembly model with a large number of experimental reaction progress traces. Using this approach, global analysis of the concentration dependence of the rate of fibril assembly over 200 reaction progress curves with 21 different possible models and subsequent ranking of the models tested suggests that nucleated amyloid assembly of β_2m under the conditions used is likely to involve the addition of monomers to form a structural nucleus approximately the size of a hexamer, despite displaying a relatively low dependence of the rate of assembly on the monomer concentration. Indeed, if the log(concentration) vs. log(t_{lag}) plot in Fig. 3 is fitted to a linear function, a slope of 0.8 results, suggesting a monomeric "critical nucleus." One reason for this apparent discrepancy is that the concentration dependence of reaction times alone provides only an estimate of the approximate size of the thermodynamic nucleus n_T, that is to say, the highest free energy species along the assembly reaction coordinate (maximum in *SI Text*, Eq. **10**), under the experimental conditions and initial monomer concentrations used. Also, for a multimolecular reaction, n_T depends on the monomer concentration, because the entropic cost of forming multimeric species varies with the initial monomer concentration, a feature that is masked when applying a standard concentration of 1 M. We define the structural nucleus size, n_S, however, using the shape of the reaction free energy for each assembly step as function of the extent of assembly (*SI Text*, Eqs. **8** and **9**) that includes terms describing the interaction pattern of the earliest species (Fig. 2*C*). Accordingly, n_S does not depend on the initial monomer concentration, and its determination thus provides important new information about the assembly process. Here, by modeling the nucleation rate and free energy functions separately from the generic polymerization reaction scheme itself, n_S (calculated from the shape of the reaction free energy as a function of the extent of assembly) (*SI Text*, Eqs. **8** and **9**) and the concentration dependence of n_T (calculated as the species with highest free energy compared with monomer) (*SI Text* Eq. **10**) are defined separately, allowing relevant structural and thermodynamic information regarding the early assembly species to be obtained.

For the *in vitro* assembly of β_2m fibrils, we show that a nucleation mechanism described by a step free energy function, involving a sudden increase in the number of stabilizing interactions leading from the structural nucleus (Fig. 2*C*), fits the data better than the other functions tested. A similar step nucleation mechanism has also been successfully used to describe assembly of other linear protein aggregates but does not describe well the aggregation of sickle-cell haemoglobin (21). This may reflect the fact that the former cases involve specific sidechain–sidechain interactions, rather than the simple packing of "large sticky spheres" as in the case of sickle cell aggregates (19). Another feature of the step nucleation mechanism is that the step free energy functions describing this type of nucleation mechanism only allow thermodynamic nucleus sizes n_T of 1 (at high concentrations), n_S, or infinity (at low concentrations where polymerization does not take place). Therefore, at intermediate protein concentrations where $n_T = n_S$, significant oligomer populations may be present in the lag-phase, relatively large structural and thermodynamic nucleus sizes will result, and shallow enough free energy functions for early species will still ensure that the rate of the overall reaction shows a small concentration dependence on the monomer concentration. This is analogous to the properties seen for a comparable polymerization mechanism through analytical approaches (12). Thus, nucleated assembly processes displaying a low dependence on the protein concentration of the rate of assembly do not necessarily exclude the possible presence of multimeric nuclei preceded by on pathway, potentially toxic, oligomers.

Role of Fragmentation in Amyloid Fibril Assembly. Mechanical agitation not only affects the rate of nucleated fibrillation reactions (8, 27), but may also affect other properties, such as the final fibril morphology (15, 27), cellular toxicity (28), and the phenotypic behavior of different prion strains (29). How agitation produces such a variety of effects, however, was not clear (14, 28). Here, we have shown that fragmentation is the dominating result of agitation in β_2m fibril formation under the conditions used, rationalizing the dramatic effect of agitation on the rate of fibril formation and the shape of the reaction progress curves.

We have shown that β_2m fibrils with different morphologies can be formed under exactly the same solution conditions, depending on the agitation condition used (15). This phenomenon has also been observed with other amyloidogenic sequences, such as A$\beta_{1\text{-}40}$ (27) and GNNQQNY from the yeast prion protein Sup35p (30). In the case of β_2m, short rod-like or worm-like fibrils are formed rapidly under some conditions without mechanical agitation (15). These species represent kinetically stable oligomers off-pathway to long-straight fibrils. If sufficient agitation is provided so that the rate of formation of long-straight fibrils is significantly accelerated through the fragmentation mechanism we propose, the assembly pathway can shift from being kinetically controlled to being thermodynamically controlled, favoring the formation of long-straight amyloid fibrils as experimentally shown. Because fibrils with different morphologies are likely to respond differently to mechanical stress (29, 31), the assembly pathway taken may be dominated by the brittleness of the fibrils and the amplitude of the agitation force. Because both the mechanical and solution conditions are critical in defining the overall assembly energy landscape, both conditions must be precisely controlled to define the structural molecular mechanism of amyloid formation, to delineate the origins of amyloid disease and to develop therapeutic agents able to control aggregation by decreasing the concentration of toxic species or the fibril load that, together, are the causative characteristics of amyloid disease.

Materials and Methods

***De Novo* Fibril Growth.** β_2m protein expression and purification were carried out as described in ref. 32. The purity of the final gel filtered samples was confirmed by SDS/PAGE, native PAGE, and MS. Lyophilized protein powder was dissolved directly into the reaction buffer containing 10 mM sodium phosphate (pH 2.0), 50 mM NaCl, 0.02% (wt/vol) NaN_3, and 10 μM ThT. The mixture was then immediately syringe-filtered (0.2 μm of MiniSart fast flow), and 100 μl was added into the wells of black 96-well plates (Corning). The plates were sealed with transparent film and put into a BMG FluorStar plate reader at 37°C to start measurement. The measurement program consisted of 1-min measurements between 5 min of orbital 1-mm diameter shaking at 600 rpm, giving a sampling frequency of 10 data points per hour. The dead time between dissolving the protein and the start of the measurement was usually 10 min, equivalent to $<$3% of the t_{lag} of even the fastest reactions. Under the mechanical agitation condition used here, we estimate that the maximum sheer stress (based on the flow speed along the wall of the wells with inner diameter of 6 mm) is several orders of magnitude lower than the sheer

modulus of the fibrils (31). However, because the agitation also induces other deformations such as torsion and bending in addition to pure sheer of the fibrils during the continuous application of agitation, significant breakage of the fibrils results. Fibril samples were analyzed by negative stain EM as described in ref. 32.

Seeded Fibril Growth. Seed solutions were made by dilution of fibril samples formed as described above. For the determination of the effect of fragmentation on seeding efficiency, the fibrils to be used as seeds were subjected to additional vigorous stirring for different lengths of time, using a 3×8 mm PTFE coated magnetic stirring bar in a 1.5-ml chromatography glass vial on a multipoint stirrer (custom made by the workshop of the School of Physics and Astronomy, University of Leeds) with an accurate rpm readout provided by a rev-counter on the rotor axis. The stirring speed was 1,500 rpm. Seeding solutions of 2% (vol/vol) were used. The reactions were initiated by adding protein monomers and fibril growth was monitored as described above, except no agitation was used. The dead time between mixing the protein solution with the seeds and start of the measurements was usually 2 min.

Data Analysis and Modeling. All data analysis and modeling calculations, including numerical integration of ordinary differential equation systems and least-square optimizations, were carried out by using Matlab (Mathworks). For the reaction progress data, scripts enabling fully automated linear baseline normalization and transition parameter extraction were used to ensure objective extraction of the curve parameters t_{lag} and k. Calculated progress curves were obtained through numerical integrations of the ordinary differential equation systems presented by the reaction kinetics. The numbers of species explicitly expressed in the calculations were chosen (up to 2400-mer), so that the error introduced by ignoring species larger than the cut-off size used was $<1\%$. t_{lag} and k values were obtained from the calculated curves as described for the experimental data and fitted globally to the experimental t_{lag} and k values. The residuals in the least-square analyses were weighted by the standard error of mean obtained from the experimental data to normalize them to the magnitude and the number of curves in the global analysis. The goodness-of-fit for tested models were evaluated by visual comparison of the fits and by comparing AIC_C scores (see *SI Text*).

ACKNOWLEDGMENTS. We thank the members of the S.E.R. and S.W.H. groups for helpful comments, Gary Thompson for helpful discussions on model comparison methods, Alison Ashcroft for stimulating discussions about MS, and Adrian Hick for technical assistance in EM. This study was funded by the Wellcome Trust.

1. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
2. Jahn TR, Radford SE (2008) Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch Biochem Biophys* 469:100–117.
3. Lesne S, Koh MT, *et al.* (2006) A specific amyloid-beta protein assembly in the brain impairs memory. *Nature* 440:352–357.
4. Ferrone FA (1999) Analysis of protein aggregation kinetics. *Methods Enzymol* 309:256–274.
5. Oosawa F, Asakura S (1975) in *Molecular biology* (Academic, London), pp 41–55.
6. Chen S, Ferrone FA, Wetzel R (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci USA* 99:11884–11889.
7. Ignatova Z, Gierasch LM (2005) Aggregation of a slow-folding mutant of a beta-clam protein proceeds through a monomeric nucleus. *Biochemistry* 44:7266–7274.
8. Nielsen L, *et al.* (2001) Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. *Biochemistry* 40:6036–6046.
9. Frankenfield KN, Powers ET, Kelly JW (2006) Influence of the N-terminal domain on the aggregation properties of the prion protein. *Protein Sci* 14:2154–2166.
10. Ellisdon AM, Pearce MC, Bottomley SP (2007) Mechanisms of ataxin-3 misfolding and fibril formation: Kinetic analysis of a disease-associated polyglutamine protein. *J Mol Biol* 368:595–605.
11. Padrick SB, Miranker AD (2002) Islet amyloid: Phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis. *Biochemistry* 41:4694–4703.
12. Powers ET, Powers DL (2006) The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the ''supercritical concentration.'' *Biophys J* 91:122–132.
13. Pellarin R, Caflisch A (2006) Interpreting the aggregation kinetics of amyloid peptides. *J Mol Biol* 360:882–892.
14. Collins SR, Douglass A, Vale RD, Weissman JS (2004) Mechanism of prion propagation: Amyloid growth occurs by monomer addition. *PLoS Biol* 2:e321.
15. Gosal WS, *et al.* (2005) Competing pathways determine fibril morphology in the self-assembly of beta2-microglobulin into amyloid. *J Mol Biol* 351:850–864.
16. Platt GW, McParland VJ, Kalverda AP, Homans SW, Radford SE (2005) Dynamics in the unfolded state of beta2-microglobulin studied by NMR. *J Mol Biol* 346:279–294.
17. Hofrichter J (1986) Kinetics of sickle hemoglobin polymerization. III. Nucleation rates determined from stochastic fluctuations in polymerization progress curves. *J Mol Biol* 189:553–571.
18. Hortschansky P, Schroeckh V, Christopeit T, Zandomeneghi G, Fändrich M (2005) The aggregation kinetics of Alzheimer's beta-amyloid peptide is controlled by stochastic nucleation. *Protein Sci* 14:1753–1759.
19. Ferrone FA (2006) Nucleation: The connections between equilibrium and kinetic behavior. *Methods Enzymol* 412:285–299.
20. Goldstein RF, Stryer L (1986) Cooperative polymerization reactions. Analytical approximations, numerical examples, and experimental strategy *Biophys J* 50:583–599.
21. Ferrone FA, Hofrichter J, Eaton WA (1985) Kinetics of sickle hemoglobin polymerization. II. A double nucleation mechanism. *J Mol Biol* 183:611–631.
22. Flyvbjerg H, Jobs E, Leibler S (1996) Kinetics of self-assembling microtubules: An ''inverse problem'' in biochemistry. *Proc Natl Acad Sci USA* 93:5975–5979.
23. Lomakin A, Teplow DB, Kirschner DA, Benedek GB (1997) Kinetic theory of fibrillogenesis of amyloid beta-protein. *Proc Natl Acad Sci USA* 94:7942–7947.
24. Bishop MF, Ferrone FA (1984) Kinetics of nucleation-controlled polymerization. A perturbation treatment for use with a secondary pathway. *Biophys J* 46:631–644.
25. Smith AM, Jahn TR, Ashcroft AE, Radford SE (2006) Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionization mass spectrometry. *J Mol Biol* 364:9–19.
26. Ruschak AM, Miranker AD (2007) Fiber-dependent amyloid formation as catalysis of an existing reaction pathway. *Proc Natl Acad Sci USA* 104:12341–12346.
27. Petkova AT, Yau WM, Tycko R (2006) Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. *Biochemistry* 45:498–512.
28. Lee S, Fernandez EJ, Good TA (2007) Role of aggregation conditions in structure, stability, and toxicity of intermediates in the Abeta fibril formation pathway. *Protein Sci* 16:723–732.
29. Tanaka M, Collins SR, Toyama BH, Weissman JS (2006) The physical basis of how prion conformations determine strain phenotypes. *Nature* 442:585–589.
30. van der Wel PC, Lewandowski JR, Griffin RG (2007) Solid-state NMR study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p. *J Am Chem Soc* 129:5117–5130.
31. Smith JF, Knowles TP, Dobson CM, Macphee CE, Welland ME (2006) Characterization of the nanoscale properties of individual amyloid fibrils. *Proc Natl Acad Sci USA* 103:15806–15811.
32. Kad NM, Thomson NH, Smith DP, Smith DA, Radford SE (2001) Beta(2)-microglobulin and its deamidated variant, N17D form amyloid fibrils with a range of morphologies in vitro *J Mol Biol* 313:559–571.

Supporting Information

Xue *et al.* 10.1073/pnas.0711664105

SI Text

Model Design. Describing the amyloid assembly process using three distinct modules enables a wide range of different assembly mechanisms to be tested systematically by combining the different variants of each module. Each possible assembly model is then composed of a variant of each of the three modules: prepolymerization, polymerization, and secondary process (Fig. 2). Below is a detailed description of each module in the analysis presented here.

Prepolymerization. The prepolymerization module describes the equilibrium between monomers at the start of the reaction, with the extension unit (X) during polymerization, or with prefibrillar species. We have chosen to test three distinct cases (Fig. 2*B*): no prepolymerization equilibrium (i.e., monomer addition), monomer-dimer equilibrium with dimer addition, and monomer conformational exchange equilibrium [i.e., a rarely populated monomeric structural nucleus as suggested by other studies of fibril assembly (1–4)]. The reaction described by simple monomer addition without other prepolymerization events does not necessarily exclude the possible presence of other events, but rather implies that the overall rate of polymerization is not altered significantly by such events under the experimental conditions and protein concentration range used. This could occur by these events having a small amplitude (the events are rare) and/or by them not being rate limiting.

Polymerization. The main polymerization module describes the reversible addition of a specific species X onto the growing oligomers or fibrils. To avoid approximations regarding populations of prenucleation species and/or species close in size to the thermodynamic nucleus, we choose to model every assembly step explicitly. The rate equation for the concentration of each species X_i is then:

$$\ldots X_{i-1} + X \underset{k_i'}{\overset{k_{i-1}}{\rightleftharpoons}} \qquad X_i + X \underset{k_{i+1}'}{\overset{k_i}{\rightleftharpoons}} X_{i+1} \ldots \qquad [1]$$

$$\frac{d[X_i]}{dt} = k_{i-1}[X_{i-1}][X] - k_i'[X_i] - k_i[X_i][X] + k_{i+1}'[X_{i+1}]$$

$$2 \le i \le N$$

where X_i are the oligomers/fibrils composed of i units of X from 2 to a large number N, k_i are the second order association rate constants, and k_i' are the first order dissociation rate constants. Because we are not constraining the model through simplifying the reaction scheme, a very large number of rate constants [2(N − 1) in total] are required to describe the full assembly of fibrils. We solve this through constraining how the free energy (and consequently the rate constants) vary with regard to the extent of assembly, as described below.

The standard reaction free energy for each assembly step as a function of the extent of polymerization, $\Delta G^\circ_{i\to(i+1)}$, can be derived from following analysis. Considering each extension step i, corresponding to the following reaction:

$$X_i + X_1 \underset{k_{i+1}'}{\overset{k_i}{\rightleftharpoons}} X_{i+1} \qquad [2]$$

the free energy change can be expressed as:

$$\Delta G_{i\to(i+1)} = \mu_{i+1} - \mu_i - \mu_1 \qquad [3]$$

where μ_i is the chemical potential of species i. μ_1 is thus the chemical potential of monomer. Several different contributions to the chemical potential can be considered for each species:

$$\begin{aligned}\mu_i &= \mu_i^\circ + RT\ln a_i \\ &= \mu^\circ_{i,\text{trans}} + \mu^\circ_{i,\text{rot}} + \mu^\circ_{i,\text{intr}} + (\mu^\circ_{i,\text{trans}} - \mu^r_{i,\text{trans}}) + RT\ln a_i\end{aligned} \qquad [4]$$

where a_i is the activity of the species i. The term $\mu G^\circ_{i,\text{intr}}$ describes the chemical potential arising from intra- and intermolecular interactions and dynamics within the species i, and contains both enthalpic and entropic contributions. The translational term $\mu G^\circ_{i,\text{trans}}$, and the rotational term $\mu G^\circ_{i,\text{rot}}$ are entropic terms related to the size of the species and can be estimated for solution species assuming ideal gas behavior (5) as:

$$\mu^\circ_{i,\text{trans}} = -RT\left[\ln\left(\frac{2\pi m k_B T}{h^2}\right)^{\frac{3}{2}} + 1 - \ln\left(\frac{N_0}{V^\circ}\right)\right] \qquad [5]$$

$$\mu^\circ_{i,\text{rot}} = -RT\ln\frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 I k_B T}{h^2}\right)^{\frac{3}{2}} \qquad [6]$$

where m is the mass, I is the moment of inertia proportional to mass times the radius squared, σ is the symmetry factor describing the shape, N_0 is the Avogadro's number, V° is the standard state volume (N_0/V° is hence the number concentration of the standard state), k_B is the Boltzmann constant, h is the Planks constant, R is the gas constant, and T is the absolute temperature. In Eq. **4**, an additional term ($\mu^\circ_{i,\text{trans}} - \mu^r_{i,\text{trans}}$) is introduced, describing the difference in translational entropy between a reference concentration (C^r) and the standard state concentration (C°). Using Eq. **5**, the difference is identified as $RT\ln(C^\circ/C^r)$, and we use the initial monomer concentration $[X]_0$ as the reference concentration here. This term is necessary to enable comparison between species involved in multimolecular reactions, which result in large changes in translational entropy due to the change in number of molecules before and after reaction, at any protein concentration (i.e., not only at the standard concentration). Assuming ideal behavior of all species, and the standard concentration of 1 M, the free energy change is then:

$$\Delta G_{i\to(i+1)} = \Delta G^\circ_{i\to(i+1)} + RT\ln[X]_0 + RT\ln\left(\frac{[X_{i+1}]}{[X_i][X_1]}\right) \qquad [7]$$

where $\Delta G^\circ_{i\to(i+1)}$ contains chemical potential terms that are size (i) dependent and a size independent part (μ°_{const}):

$$\begin{aligned}\Delta G^\circ_{i\to(i+1)} &= \mu^\circ_{(i+1)} - \mu^\circ_i - \mu^\circ_1 \\ &= (\mu^\circ_{(i+1),trans} + \mu^\circ_{(i+1),\text{rot}} + \mu^\circ_{(i+1),\text{intr}} \\ &\quad - (\mu^\circ_{i,trans} + \mu^\circ_{i,\text{rot}} + \mu^\circ_{i,intr}) - \mu^\circ_1 \\ &= (\mu^\circ_{(i+1),trans} - \mu^\circ_{i,\text{trans}}) + (\mu^\circ_{(i+1),rot} - \mu^\circ_{i,\text{rot}}) \\ &\quad + (\mu^\circ_{(i+1),\text{intr}} - \mu^\circ_{i,\text{intr}}) - \mu^\circ_1 \\ &= \Delta\mu^\circ_{i\to(i+1),\text{trans}}(i) + \Delta\mu^\circ_{i\to(i+1),\text{rot}}(i) \\ &\quad + \Delta\mu^\circ_{i\to(i+1),\text{intr}}(i) + \mu^\circ_{\text{const}}\end{aligned} \qquad [8]$$

At equilibrium, $\Delta G_{i\rightarrow(i+1)}$ is zero. This yields the standard reaction free energy $\Delta G^{\circ}_{i\rightarrow(i+1)}$ for each extension step as function of the concentration of each species and the initial monomer concentration $[X]_0$:

$$\Delta G^{\circ}_{i\rightarrow(i+1)} = -RT\ln\left(\frac{[X_{i+1}]}{[X_i][X_1]}\right) - RT\ln[X]_0 = -RT\ln\left(\frac{k_i}{k'_{i+1}}\right) - RT\ln[X]_0 \quad [9]$$

The free energy difference between any species i and monomers is then determined as:

$$\Delta G^{\circ}_{1\rightarrow i} = \sum_{j=1}^{i-1} \Delta G^{\circ}_{j\rightarrow(j+1)} = \sum_{j=1}^{i-1}\left[-RT\ln\left(\frac{k_j}{k'_{j+1}}\right)\right] - (i-1)RT\ln[X]_0 \quad [10]$$

The thermodynamic nucleus can now be defined as the species with $i = n_T$ that has the highest $\Delta G^{\circ}_{1\rightarrow(i+1)}$ (i.e., the highest free energy compared with the monomer), and this species controls the overall rate of the reaction. In Eq. **10**, the $-(i-1)\mathrm{RTln}[X]_0$ term corrects for the (unfavorable) increase in free energy due to the loss of the translational entropy caused by the reduction in the number of particles during assembly at different initial protein concentrations, a phenomenon that makes the thermodynamic nucleus size n_T concentration dependent.

From the above analysis, the overall behavior of the assembly reaction can now be attributed to the interplay between the initial monomer concentration, and the size dependence of the reaction free energy. In Eq. **8**, we identify the size dependent contributions as the changes in translational $\Delta\mu^{\circ}_{i\rightarrow(i+1),\mathrm{trans}}$ and rotational $\Delta\mu^{\circ}_{i\rightarrow(i+1),\mathrm{rot}}$ entropies and the changes in interaction free energies for each step $\Delta\mu^{\circ}_{i\rightarrow(i+1),\mathrm{intr}}$. Using Eqs. **5** and **6**, the changes in translational and rotational entropies for each addition step are:

$$\Delta\mu^{\circ}_{i\rightarrow(i+1),\mathrm{trans}}(i) = -RT\ln\left(\frac{i+1}{i}\right)^{\frac{3}{2}} \quad [11]$$

$$\Delta\mu^{\circ}_{i\rightarrow(i+1),\mathrm{rot}}(i) = -RT\ln\left(\frac{i+1}{i}\right)^{\frac{5}{2}} \quad [12]$$

From the above relations, we can see that the size dependent translational and rotational entropy changes are greatest at the beginning of assembly as expected. The largest free energy changes are obtained in the monomer to dimer step, resulting in a doubling in mass. However, this only results in -2.7 kJmol^{-1} in translational entropy change and -4.5 kJmol^{-1} in rotational entropy change (assuming the shape stays the same) at 310 K, using Eqs. **11** and **12**. Thus, it is possible for the interaction term in Eq. **8** that includes structural information to dominate the overall size dependence of the reaction free energy for amyloid fibril assemblies, which contain a significant number of specific interactions and structural symmetry as judged from available structural information (6). Furthermore, the translational and rotational contributions are estimated above by using the ideal gas approximation. The magnitude of the translational and rotational entropy changes in solution is likely to be significantly reduced compared with the estimated values in ideal gases, possibly placing further emphasizes on the interaction terms. If information is available on the shape of the reaction free energy function $\Delta G^{\circ}_{i\rightarrow(i+1)}$, such information can be used to constrain the model. Since such information is usually not known, however, we approach this issue from the reverse by defining different possible shapes for the $\Delta G^{\circ}_{i\rightarrow(i+1)}$ function for the polymerization module and determining which of the functions tested allow the best-fit of the assembly model to the experimental data collected, over the entire concentration range of the rates measured. Using our modular modeling approach, we are hence able to test different functional shapes of $\Delta G^{\circ}_{i\rightarrow(i+1)}$ that correspond to diverse possibilities (Fig. 2*C*).

We selected three different functional shapes of $\Delta G^{\circ}_{i\rightarrow(i+1)}$ to test to cover different possible scenarios. Fig. 2*C* in shows the three different free energy functions tested in this study: step, linear or power functions, together with simple diagrams illustrating the interaction patterns corresponding to the different free energy functions. The step free energy profile represents a nucleation mechanism involving a sudden change in the number of stabilizing interactions leading to and from the structural nucleus (Fig. 2*C Top)*. This type of step nucleation mechanism has been used to describe other nucleated polymerization processes, such as for actin and flagellin (7–10), and places an emphasis on the creation of additional specific interactions when the species reach a certain size. The power free energy function tested in this study represents a gradual change of the reaction free energy as size of the species increase (Fig. 2*C Bottom*), analogous to mechanisms used to describe nucleated phase transitions and similar models applied to the aggregation of sickle cell hemoglobin (5, 11, 12). By using a power function, this free energy profile mimics the size dependence of the reaction free energy to geometric considerations when packing spheres (13) and to a situation that places emphasis on translational and rotational considerations. The linear nucleation profile (Fig. 2*C Middle*) is an intermediate between the step and the power free energy profiles and represents a gradual increase in the number of stabilizing interactions during nucleation.

In summary, because of the significant loss of translational entropy during reactions associated with multimolecular assembly such as fibril formation, if the experiments are carried out at a high enough initial monomer concentration, n_T will become one (denoting a down hill assembly reaction in the thermodynamic sense, Eq. **10**) independent of any other considerations and this trivial behavior does not carry any structural significance by itself. By using the approach outlined above, we effectively reference the free energies to the working monomer concentration, allowing the difference in free energy between different polymerization species to be determined at any initial monomer concentration, while still retaining the use of standard reaction free energies referenced at 1 M. An important feature of our approach is that the reaction scheme defined in (1) does not impose any additional constraints, for example whether the overall process is following a nucleated mechanism, with the only assumption being that the addition of only a single species can occur. Instead, the choice of the rate and free energy functions, k_i, k'_i and G°_i impose such constraints, enabling different assembly possibilities to be tested. We can now introduce the structural nucleus size (n_S), defined by the shape of the free energy function in Eq. **9**. Because the free energy function itself, being corrected for concentration dependent terms, only has size dependent terms and constant contributions, n_S is independent of protein concentration. This parameter therefore can provide useful information on the interactions involved in early species, linking the thermodynamic aspects to the structural aspects of assembly. This approach to modeling the polymerization process through rate and free energy functions hence enables the thermodynamic and kinetic properties observed in experiments to be linked to possible structural aspects of polymerization by

having separate and exact definitions for the structural nucleus size n_S and the thermodynamic nucleus size n_T.

Fragmentation. The third module considers secondary processes and allows feedback responses that can modulate polymerization to be considered. Possible secondary effects include heterogeneous nucleation on fibril surfaces, branching or fragmentation of fibrils (14). Heterogeneous nucleation has been shown to be involved in the assembly of sickle cell-hemoglobin (11, 12) and some amyloid assembly mechanisms (15, 16), while in other cases fragmentation has been suggested as the principle secondary process (17, 18). Mechanical agitation has profound effects on the rate of fibrillation of β_2m and other proteins, which suggests that fragmentation may be a dominating secondary process for fibril assembly, especially under conditions such as those used here in which agitation is required for fibril formation to occur in a readily measurable time scale. Here, fragmentation is assumed to be an irreversible process, cleaving X_i into two fragments X_j and X_k (Fig. 2*D*). The irreversibility of fragmentation reasonably assumes that the rate of mechanically forced fragmentation due to agitation is much greater than the back reaction of fibril reannealing. The first order fragmentation rate constant can then be expressed as a function of the size of the fragmenting fibril species (i) and the location of the fragmentation site (j) according to following equation based on statistical mechanical considerations for linear polymers (19):

$$k_{Fr}(i,j) = c_1[j(i-j)]^{c_2-1}\left[\frac{(i-j)lnj + j\ln(i-j)}{i^{c_2+1}}\right] \quad [13]$$

In the above equation, c_1 describes the overall amplitude and c_2 describes the size and position dependence of the fragmentation rate constant. This approach in which both the length and positional dependence of fragmentation are considered, builds on simpler methods (17), but still assumes that fragmentation can occur between any two units in a fibril. More specific fragmentation events that could occur, for example structural mechanical consideration leading to alternative, more specific, fragmentation patterns, are not considered. More detailed information on the structural-mechanical properties of fibrils will be needed to justify more complex alternatives.

Progress Curves. Using the above modular approach, the progress curves for different possible assembly mechanisms are calculated numerically. This eliminates the need for additional, potentially erroneous, assumptions commonly used in analytical models, such as negligible oligomer populations and irreversible extensions (10, 14). The primary output of the model is the distribution of species at any given time during the reaction. To compare the model with the experimental data, we assume that the fluorescence of thioflavin T responds linearly to the mass fraction of monomers present in the fibrils [X_F]. A fibril is assumed to be any species containing more monomers than a structural nucleus (i.e., $i \geq n_S + 1$) (10) giving the following expression for the calculating progress curves as function of time t:

$$[X_F](t) = \sum_{i=n_S+1}^{N} i[X_i](t) \quad [14]$$

Model Comparison. From the numerically calculated progress curves, the t_{lag}, and k values are obtained from in the same way as for the experimental data and fitted globally to the experimental data (see *Materials and Methods* in the main text for details regarding numerical calculations and weighting). The goodness of fit for tested models are evaluated by visual comparison of the fits and by comparing corrected AIC (AIC_c) scores (20), calculated based on the number of data points N, the number of fitted parameters M, and the residual sum of squares RSS according to:

$$AIC_C = N\ln\left(\frac{RSS}{N}\right) + 2M + \frac{2M(M+1)}{N-M-1} \quad [15]$$

The probability of a model being more accurate compared with the best-fit model (with lowest AIC_C score) despite a less good fit is calculated as:

$$\text{Probability} = \frac{e^{-0.5(AIC_c - AIC_{c,best-fit})}}{1 + e^{-0.5(AIC_c - AIC_{c,best-fit})}} \quad [16]$$

1. Chen S, Ferrone FA, Wetzel R (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci USA* 99:11884–11889.
2. Thakur AK, Wetzel R (2002) Mutational analysis of the structural organization of polyglutamine aggregates. *Proc Natl Acad Sci USA* 99:17014–17019.
3. Bhattacharyya AM, Thakur AK, Wetzel R (2005) Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction. *Proc Natl Acad Sci USA* 102:15400–15405.
4. Bhattacharyya A, *et al.* (2006) Oligoproline effects on polyglutamine conformation and aggregation. *J Mol Biol* 355:524–535.
5. Ferrone FA (2006) Nucleation: The connections between equilibrium and kinetic behavior. *Methods Enzymol* 412:285–299.
6. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
7. Wakabayashi K, Hotani H, Asakura S (1969) Polymerization of Salmonella flagellin in the presence of high concentrations of salts. *Biochim Biophys Acta* 175:195–203.
8. Frieden C, Goddette DW (1983) Polymerization of actin and actin-like systems: Evaluation of the time course of polymerization in relation to the mechanism. *Biochemistry* 22:5836–5843.
9. Sept D, McCammon JA (2001) Thermodynamics and kinetics of actin filament nucleation. *Biophys J* 81:667–674.
10. Powers ET, Powers DL (2006) The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the ''supercritical concentration.'' *Biophys J* 91:122–132.
11. Ferrone FA, Hofrichter J, Eaton WA (1985) Kinetics of sickle hemoglobin polymerization. I. Studies using temperature-jump and laser photolysis techniques. *J Mol Biol* 183:591–610.
12. Ferrone FA, Hofrichter J, Eaton WA (1985) Kinetics of sickle hemoglobin polymerization. II. A double nucleation mechanism. *J Mol Biol* 183:611–631.
13. Ginell R (1961) Geometric basis of phase change. *J Chem Phys* 34:992–998.
14. Ferrone FA (1999) Analysis of protein aggregation kinetics. *Methods Enzymol* 309:256–274.
15. Padrick SB, Miranker AD (2002) Islet amyloid: Phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis. *Biochemistry* 41:4694–4703.
16. Librizzi F, Rischel C (2005) The kinetic behavior of insulin fibrillation is determined by heterogeneous nucleation pathways. *Protein Sci* 14:3129–3134.
17. Collins SR, Douglass A, Vale RD, Weissman JS (2004) Mechanism of prion propagation: Amyloid growth occurs by monomer addition. *PLoS Biol* 2:e321.
18. Tanaka M, Collins SR, Toyama BH, Weissman JS (2006) The physical basis of how prion conformations determine strain phenotypes. *Nature* 442:585–589.
19. Hill TL (1983) Length dependence of rate constants for end-to-end association and dissociation of equilibrium linear aggregates *Biophys J* 44:285–288.
20. Hurvich CM, Tsai CL (1989) Regression and time-series model selection in small samples. *Biometrika* 76:297–307.

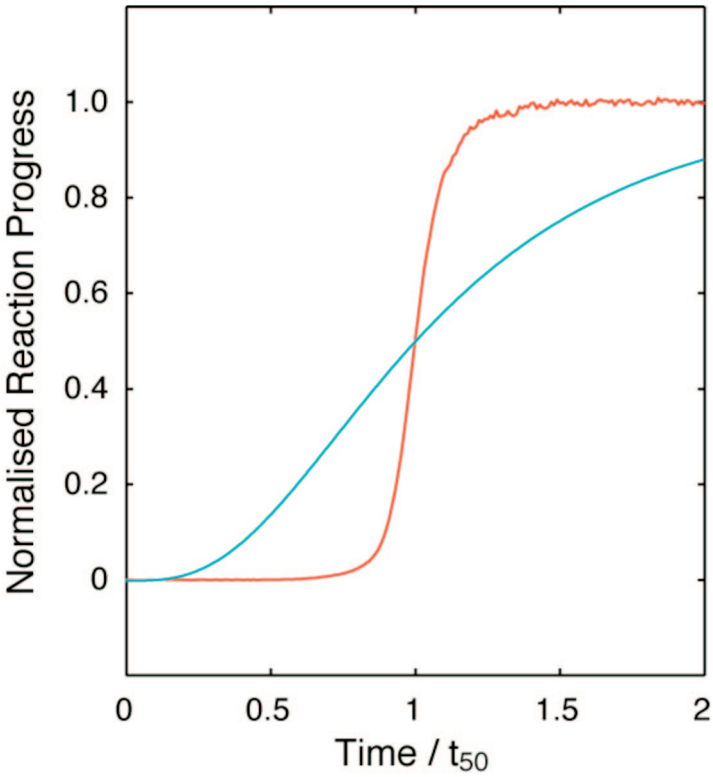

Fig. S1. Comparison of the progress curve predicted by a nucleated assembly model alone (cyan) (14) with a typical reaction progress curve obtained experimentally for β_2m (red). The progress curves are normalized to t_{50} in the *x* axis and the reaction progress at equilibrium in the *y* axis to facilitate comparison.

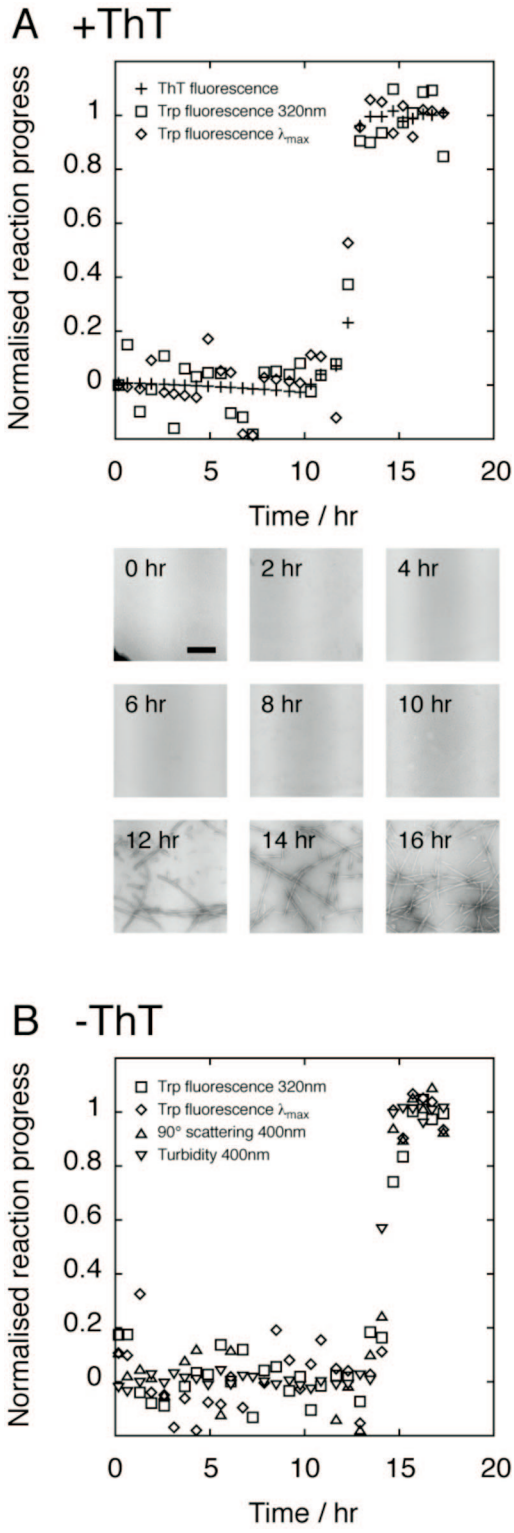

Fig. S2. Experiments monitoring the rate of fibril formation using different modes of detection. The fibril growth of two samples at the same protein concentration, one containing ThT (*A*) and one without ThT (*B*) were monitored. (*A*) Fibril growth of a sample containing ThT was monitored using ThT fluorescence (+), intrinsic Trp fluorescence (excitation at 295 nm; square, emission at 320 nm; diamond, emission λ_{max}), and negative stain EM (bar is 200 nm, note that amorphous aggregates are not observed in the lag time under these experimental conditions). (*B*) Fibril growth of a sample without ThT monitored by intrinsic Trp fluorescence (excitation at 295 nm; square: emission at 320 nm, and diamond: emission λ_{max}), 90° light scattering at 400 nm (upward triangle), and turbidity at 400 nm (downward triangle).

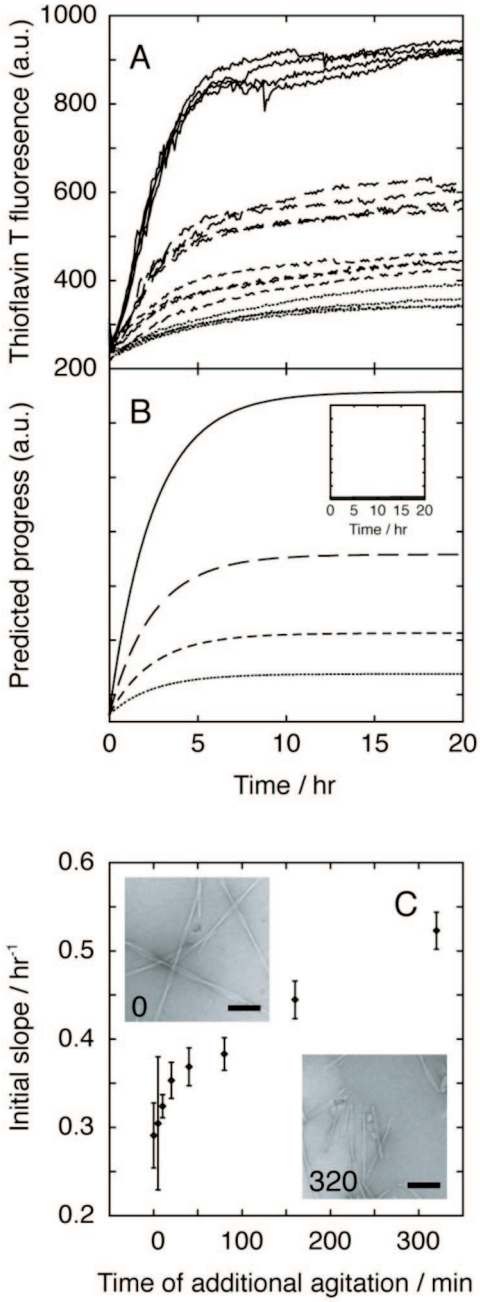

Fig. S3. Comparison of predictions made by the best-fit model and experimental data from seeded reactions. (*A*) Experimental reaction progress curves at initial monomer concentrations of 12 μM (solid lines), 6 μM (long dashed lines), 3 μM (short dashed lines), and 1.5 μM (dotted lines). Four replicates were obtained at each monomer concentration. (*B*) Reaction progress curves predicted using the best-fit model involving monomer addition, step nucleation polymerization with $n_S = 6$, and secondary fragmentation at the same monomer concentrations as in *A*. (*Inset*) Reaction progress curves predicted using the same model but without fragmentation. (*C*) Seed efficiency of fibrils fragmented to different extents. The initial slope of normalized progress curves (quiescent growth) is plotted vs. the extent of fragmentation of the fibril seeds defined by the time of agitation before their addition to the growth assays. Error bars represent one standard deviation obtained from four replicates at each data point. EM images show the fibril seeds before and after 320 min of agitation. (Scale bars, 100 nm.)

Table S1. Reaction scheme and rate equations for all tested possibilities of prepolymerization, polymerization, and secondary process

Module	Variant	Reaction scheme*	Rate equations	Rate constants†
Prepolymerization	No prepolymerization	$X = X_1$	$\frac{d[X_1]}{dt} = \begin{cases} -2k_1[X]^2 + 2k'_2[X_2] \\ -\sum_{i=2}^{N} k_i[X_i][X] + \sum_{i=2}^{N} k'_{i+1}[X_{i+1}] \end{cases}$	—
	Monomer–dimer equilibrium and dimer addition	$X_M + X_M \underset{k'_p}{\overset{k_p}{\rightleftharpoons}} X_1$ $X = X_1$	$\frac{d[X_1]}{dt} = \begin{cases} -2k_1[X_1]^2 + 2k'_2[X_2] \\ -\sum_{i=2}^{N} k_i[X_i][X] + \sum_{i=2}^{N} k'_{i+1}[X_{i+1}] \\ +k_p[X_M]^2 - k'_p[X_1] \end{cases}$ $\frac{d[X_M]}{dt} = -2k_p[X_M]^2 + 2k'_p[X_1]$	$k_p = c_{\text{pre},1}$ and $k'_p = c_{\text{pre},2}$
	Monomer conformational exchange	$X \underset{k'_p}{\overset{k_p}{\rightleftharpoons}} X_1$	$\frac{d[X]}{dt} = \begin{cases} -k_1[X_1][X] + k'_2[X_2] \\ -\sum_{i=2}^{N} k_i[X_i][X] + \sum_{i=2}^{N} k'_{i+1}[X_{i+1}] \\ -k_p[X] + k'_p[X_1] \end{cases}$ $\frac{d[X_1]}{dt} = k_p[X] - k'_p[X_1] - k_1[X_1][X] + k'_2[X_2]$	$k_p = c_{\text{pre},1}$ and $k'_p = c_{\text{pre},2}$
Polymerization	Step rate and free energy function Linear nucleation rate and free energy function	$X_1 \underset{k'_2}{\overset{k_1}{\rightleftharpoons}} X_2 \underset{k'_3}{\overset{k_2}{\rightleftharpoons}} X_3 \ldots$	$\frac{d[X_i]}{dt} = \begin{cases} k_{i-1}[X_{i-1}][X] - k'_i[X_i] \\ -k_i[X_i][X] + k'_{i+1}[X_{i+1}] \end{cases} \quad 2 \le i \le N$	$k_i = c_1;\quad k'_{i+1} = c_3;\quad i < n_S$ $k_i = c_2;\quad k'_{i+1} = c_4;\quad i \ge n_S$ $\left.\begin{array}{l} \log k_i = c_1 + \left(\frac{c_2 - c_1}{n-1}\right)(i-1) \\ \log k'_{i+1} = c_3 + \left(\frac{c_4 - c_3}{n-1}\right)(i-1) \end{array}\right\} \; i < n_s$ $\log k_{i+1} = c_2 \log k'_{i+1} = c_4 \quad i \ge n_s$
	Power rate and free energy function			$\log k_i = c_1 i^{c5} + c_2;$ $\log k'_i = c_3 i^{c6} + c_4$
Secondary process	No secondary process	—	$\left[\frac{d[X_i]}{dt}\right]_{2\text{nd}} = 0$	—
	Fragmentation‡	$X_i \xrightarrow{k_{Fr}(i,j)} X_j + X_k$ $i = j + k$	$\left[\frac{d[X_i]}{dt}\right]_{2\text{nd}=Fr} = -\sum_{j=1}^{i-1} k_{Fr}(i,j)[X_i] + 2\sum_{j=i+1}^{N} k_{Fr}(j,i)[X_j]$	$k_{Fr}(i,j) = c_{2\text{nd},1}[j(i-j)]^{c_{2\text{nd},2}-1}\left[\frac{(i-j)\ln j + j\ln(i-j)}{i^{c_{2\text{nd},2}+1}}\right]$

*X denotes the elongating species; X_M are the prepolymerization species.
†Parameters are denoted as c.
‡Rate constants derived from statistical mechanical treatment of a linear polymer (19). Reaction is assumed to be irreversible because the reverse reaction is much slower than forced fragmentation.

Table S2. Summary of all tested mechanisms and their AIC_C scores

Model no.	Prepolymerization*	Polymerization	Secondary process	n_S	Fitted parameters, no.	AIC_C-$AIC_{C\ best\text{-}fit}$[†]	Probability, %[‡]
1	None	Step	No fragmentation	6	4	87.6	<0.1
2	None	Step	Fragmentation	1	4	132.7	<0.1
3	None	Step	Fragmentation	2	6	9.7	0.8
4	None	Step	Fragmentation	4	6	20.9	<0.1
5	None	Step	Fragmentation	6	6	0.0	Best-fit
6	None	Step	Fragmentation	8	6	39.6	<0.1
7	None	Step	Fragmentation	10	6	54.4	<0.1
8	None	Linear	Fragmentation	4	6	1.3	34.1
9	None	Linear	Fragmentation	6	6	4.9	7.8
10	None	Linear	Fragmentation	8	6	5.6	5.8
11	None	Linear	Fragmentation	10	6	9.8	0.7
12	None	Power	No fragmentation	10[§]	6	63.7	<0.1
13	None	Power	Fragmentation	2[§]	8	9.2	1.0
14	Monomer–dimer	Step	Fragmentation	1 dimer[¶]	6	134.8	<0.1
15	Monomer–dimer	Step	Fragmentation	2 dimers[¶]	8	23.0	<0.1
16	Monomer–dimer	Step	Fragmentation	3 dimers[¶]	8	22.0	<0.1
17	Monomer–dimer	Step	Fragmentation	4 dimers[¶]	8	14.2	0.1
18	Monomer–dimer	Step	Fragmentation	5 dimers[¶]	8	18.5	<0.1
19	Monomer–dimer	Power	Fragmentation	2 dimers[§¶]	10	99.8	<0.1
20	Conf. exchange	Step	No fragmentation	1	4	126.2	<0.1
21	Conf. exchange	Step	Fragmentation	1	6	1.0	37.5

*No prepolymerization is equivalent to simple monomer addition.

[†]The absolute difference between AIC_C scores of the tested models and the best-fit model. Lower numbers indicates a better fit.

[‡]Probability that model is correct despite less good fit in comparison with the best-fit model.

[§]Free energy profiles defined by a power function do not have specific n_S. The number in the table is an estimation based on the fitted free energy profile using the method explained in Fig 2*C*.

[¶]The size n_S in monomer unit is two times n_S in dimer units (listed in the table) with the Monomer–dimer prepolymerization equilibrium.