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:

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

$$\frac{d[X_i]}{dt} = k_{i-1}[X_{i-1}][X] - k_i'[X_i] - k_i[X_i][X] + k_{i+1}'[X_{i+1}] \qquad 2 \le i \le N$$

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:

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

where a_i is the activity of the species i. The term $\mu G^\circ_{i,\mathrm{intr}}$ 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 $\mu G^\circ_{i,\mathrm{trans}}$, and the rotational term $\mu G^\circ_{i,\mathrm{rot}}$ are entropic terms related to the size of the species and can be estimated for solution species assuming ideal gas behavior (5) as:

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

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

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 ($\mu^\circ_{i,\mathrm{trans}} - \mu^r_{i,\mathrm{trans}}$) 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 ($\mu^\circ_{\mathrm{const}}$):

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

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

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

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

$$\begin{aligned}\Delta G^{\circ}_{1\to i} &= \sum_{j=1}^{i-1} \Delta G^{\circ}_{j\to(j+1)} \\ &= \sum_{j=1}^{i-1}\left[-RT\ln\left(\frac{k_j}{k'_{j+1}}\right)\right] - (i-1)RT\ln[X]_0\end{aligned} \tag{10}$$

The thermodynamic nucleus can now be defined as the species with $i = n_{\mathrm{T}}$ that has the highest $\Delta G^{\circ}_{1\to(i+1)}$ (i.e., the highest free energy compared with the monomer), and this species controls the overall rate of the reaction. In Eq. **10**, the $-(i-1)\mathrm{RT}\ln[X]_0$ 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 $\Delta\mu^{\circ}_{i\to(i+1),\mathrm{trans}}$ and rotational $\Delta\mu^{\circ}_{i\to(i+1),\mathrm{rot}}$ entropies and the changes in interaction free energies for each step $\Delta\mu^{\circ}_{i\to(i+1),\mathrm{intr}}$. Using Eqs. **5** and **6**, the changes in translational and rotational entropies for each addition step are:

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

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

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 $-2.7\ \mathrm{kJmol}^{-1}$ in translational entropy change and $-4.5\ \mathrm{kJmol}^{-1}$ 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 $\Delta G^{\circ}_{i\to(i+1)}$, 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 $\Delta G^{\circ}_{i\to(i+1)}$ 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 $\Delta G^{\circ}_{i\to(i+1)}$ that correspond to diverse possibilities (Fig. 2*C*).

We selected three different functional shapes of $\Delta G^{\circ}_{i\to(i+1)}$ 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:

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

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.

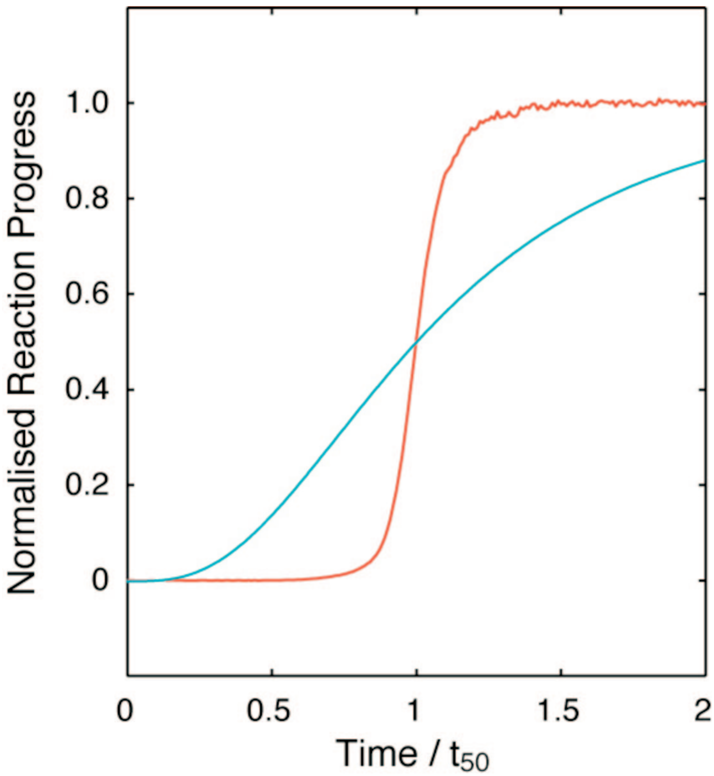

Fig. S1. Comparison of the progress curve predicted by a nucleated assembly model alone (cyan) (14) with a typical reaction progress curve obtained experimentally for β_2m (red). The progress curves are normalized to t_{50} in the x axis and the reaction progress at equilibrium in the y axis to facilitate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

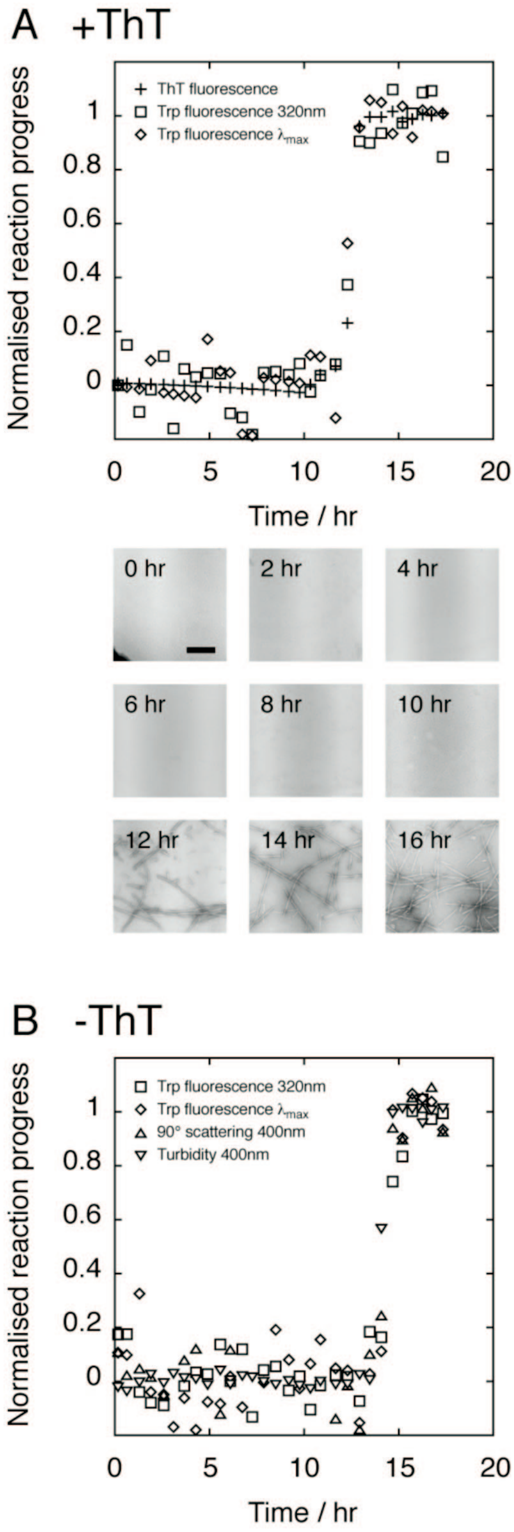

Fig. S2. Experiments monitoring the rate of fibril formation using different modes of detection. The fibril growth of two samples at the same protein concentration, one containing ThT (*A*) and one without ThT (*B*) were monitored. (*A*) Fibril growth of a sample containing ThT was monitored using ThT 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).

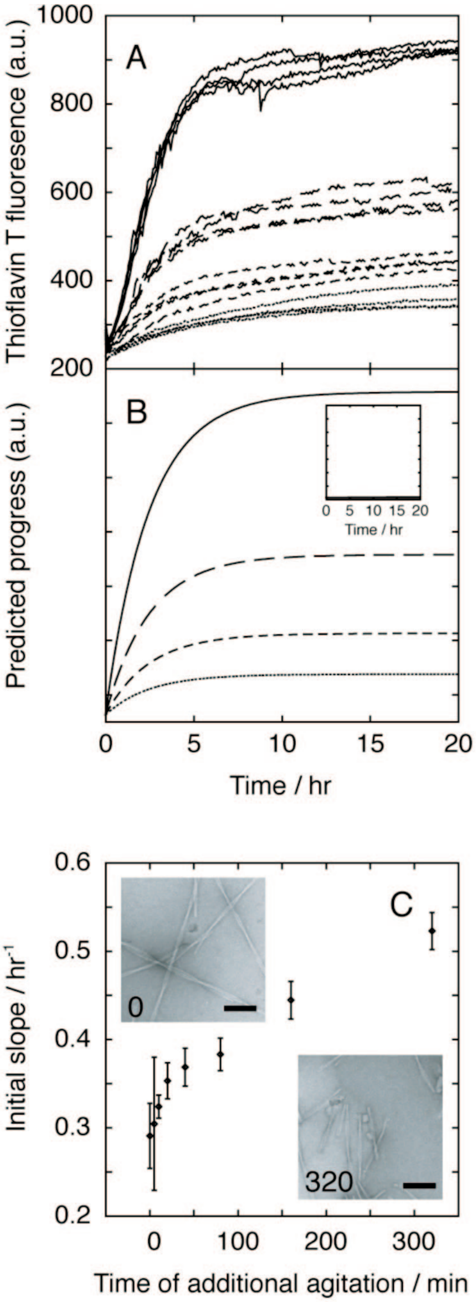

Fig. S3. Comparison of predictions made by the best-fit model and experimental data from seeded reactions. (*A*) Experimental reaction progress curves at initial monomer concentrations of 12 μM (solid lines), 6 μM (long dashed lines), 3 μM (short dashed lines), and 1.5 μM (dotted lines). Four replicates 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	$X = X_1$	$\frac{d[X_1]}{dt} = \left\{ \begin{array}{l} -2k_1[X]^2 + 2k_2'[X_2] \\ -\sum_{i=2}^{N} k_i[X_i][X] + \sum_{i=2}^{N} k_{i+1}'[X_{i+1}] \end{array} \right.$	—
	Monomer–dimer equilibrium and dimer addition	$X_M + X_M \underset{k_p'}{\overset{k_p}{\rightleftarrows}} X_1$ $X = X_1$	$\frac{d[X_1]}{dt} = \left\{ \begin{array}{l} -2k_1[X_1]^2 + 2k_2'[X_2] \\ -\sum_{i=2}^{N} k_i[X_i][X] + \sum_{i=2}^{N} k_{i+1}'[X_{i+1}] \\ + k_p[X_M]^2 - k_p'[X_1] \end{array} \right.$ $\frac{d[X_M]}{dt} = -2k_p[X_M]^2 + 2k_p'[X_1]$	$k_p = c_{\mathrm{pre},1}$ and $k_p' = c_{\mathrm{pre},2}$
	Monomer conformational exchange	$X \underset{k_p'}{\overset{k_p}{\rightleftarrows}} X_1$	$\frac{d[X]}{dt} = \left\{ \begin{array}{l} -k_1[X_1][X] + k_2'[X_2] \\ -\sum_{i=2}^{N} k_i[X_i][X] + \sum_{i=2}^{N} k_{i+1}'[X_{i+1}] \\ -k_p[X] + k'_p[X_1] \end{array} \right.$ $\frac{d[X_1]}{dt} = k_p[X] - k'_p[X_1] - k_1[X_1][X] + k_2'[X_2]$	$k_p = c_{\mathrm{pre},1}$ and $k_p' = c_{\mathrm{pre},2}$
Polymerization	Step rate and free energy function	$\begin{array}{ccccc} & X & & X & \\ & \downarrow & & \downarrow & \\ & k_1 & & k_2 & \\ X_1 & \rightleftarrows & X_2 & \rightleftarrows & X_3\ldots \\ & k_2' & & k_3' & \\ & \downarrow & & \downarrow & \\ & X & & X & \end{array}$	$\frac{d[X_i]}{dt} = \left\{ \begin{array}{l} k_{i-1}[X_{i-1}][X] - k_i'[X_i] \\ -k_i[X_i][X] + k_{i+1}'[X_{i+1}] \end{array} \right. \quad 2 \le i \le N$	$k_i = c_1; \quad k_{i+1}' = c_3; \quad i < n_S;$ $k_i = c_2; \quad k_{i+1}' = c_4; \quad i \ge n_S$
	Linear nucleation rate and free energy function			$\left. \begin{array}{r} \log k_i = c_1 + \left(\frac{c_2 - c_1}{n-1}\right)(i-1) \\ \log k_{i+1}' = c_3 + \left(\frac{c_4 - c_3}{n-1}\right)(i-1) \end{array} \right\} \quad i < n_s$ $\log k_{i+1} = c_2 \log k_{i+1}' = c_4 \quad i \ge n_s$
	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	—	$\left[\frac{d[X_i]}{dt}\right]_{\mathrm{2nd}} = 0$	—
	Fragmentation[‡]	$X_i \xrightarrow{k_{Fr}(i,j)} X_j + X_k$ $i = j + k$	$\left[\frac{d[X_i]}{dt}\right]_{\mathrm{2nd}=Fr} = -\sum_{j=1}^{i-1} k_{Fr}(i,j)[X_i] + 2\sum_{j=i+1}^{\mathrm{N}} \mathrm{k_{Fr}(j,i)[X_j]}$	$k_{Fr}(i,j) = c_{\mathrm{2nd},1}[j(i-j)]^{c_{\mathrm{2nd},2}-1}\left[\frac{(i-j)\ln j + j\ln(i-j)}{i^{c_{\mathrm{2nd},2}+1}}\right]$

---

*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.