Article

Competition between Primary Nucleation and Autocatalysis in Amyloid Fibril Self-Assembly

Kym Eden,[1,*] Ryan Morris,[1] Jay Gillam,[1] Cait E. MacPhee,[1] and Rosalind J. Allen[1]
[1]School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom

ABSTRACT Kinetic measurements of the self-assembly of proteins into amyloid fibrils are often used to make inferences about molecular mechanisms. In particular, the lag time—the quiescent period before aggregates are detected—is often found to scale with the protein concentration as a power law, whose exponent has been used to infer the presence or absence of autocatalytic growth processes such as fibril fragmentation. Here we show that experimental data for lag time versus protein concentration can show signs of kinks: clear changes in scaling exponent, indicating changes in the dominant molecular mechanism determining the lag time. Classical models for the kinetics of fibril assembly suggest that at least two mechanisms are at play during the lag time: primary nucleation and autocatalytic growth. Using computer simulations and theoretical calculations, we investigate whether the competition between these two processes can account for the kinks which we observe in our and others' experimental data. We derive theoretical conditions for the crossover between nucleation-dominated and growth-dominated regimes, and analyze their dependence on system volume and autocatalysis mechanism. Comparing these predictions to the data, we find that the experimentally observed kinks cannot be explained by a simple crossover between nucleation-dominated and autocatalytic growth regimes. Our results show that existing kinetic models fail to explain detailed features of lag time versus concentration curves, suggesting that new mechanistic understanding is needed. More broadly, our work demonstrates that care is needed in interpreting lag-time scaling exponents from protein assembly data.

INTRODUCTION

Amyloid fibrils are structured polymeric aggregates of protein molecules, which form when proteins misfold, such that they stack together in a cross-β-sheet conformation. Understanding the mechanisms involved in the self-assembly of these fibrils is of great importance, both because they are implicated in many degenerative diseases (1), and because they have potential applications in the design of new materials (2,3). The molecular processes involved in the very early stages of aggregation are of particular interest because increasing evidence suggests that early aggregates, rather than mature fibrils, may be the toxic species in fibril-linked diseases (4–6). In this article, we investigate the interplay between two competing processes, primary nucleation and autocatalytic growth, during the early stages of amyloid fibril aggregation, and assess whether this interplay can account for so-far unexplained features of our own and others' experimental data.

Amyloid fibril formation is commonly studied by monitoring the kinetics of self-assembly in vitro. In these experiments, a protein sample is placed under conditions where fibril formation is favored, and measurements of fluorescence (with β-sheet binding fluorescent dyes) or absorbance are made as a function of time. The resulting kinetic curves typically show a sigmoidal shape, with an initial lag time in which little fibrillar material is detected, followed by a period of rapid growth, and finally saturation as the pool of available unaggregated protein is exhausted. Addition of preformed fibril seeds at the start of the experiment usually abolishes the lag phase, suggesting that aggregation is initiated by a nucleation process, in which a rare fluctuation leads to the energetically unfavorable formation of the smallest stable growth-competent aggregate (primary nucleus). Once a fibril has been formed, it is expected to grow by sequential addition of protein molecules at its ends (and possibly by end-joining with other fibrils (7)).

Fitting observed kinetic data to theoretical models provides a powerful tool for elucidating the molecular mechanisms involved in fibril self-assembly (8,9). This approach has shown that the sigmoidal shape of typical kinetic curves cannot be explained by models that only involve primary nucleation and fibril growth; it also requires autocatalytic formation of new fibrils from existing ones (8,10–13). The breaking of fibrils into shorter fragments (fragmentation) provides one such autocatalysis mechanism, because it increases the number of growth-competent ends. Alternatively, or additionally, autocatalysis could happen via the nucleation of new fibrils on the surfaces of existing ones (secondary nucleation) (11–13).

The results of in vitro kinetic experiments are often summarized by plotting the measured lag time τ_{lag} as a function of the protein concentration m_{tot}, defining τ_{lag} as the time at

Submitted January 13, 2014, and accepted for publication November 26, 2014.

*Correspondence: eden.kym@gmail.com

Editor: Elizabeth Rhoades.

0006-3495/15/02/0632/12 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.11.3465

which the concentration of aggregated protein reaches a predefined threshold (although other definitions are also used (11,14–16)). The lag time decreases as the protein concentration increases, and this relation is typically fitted to a power-law function $\tau_{\mathrm{lag}} \propto {m_{\mathrm{tot}}}^{-\gamma}$ (i.e., one plots $\log(\tau_{\mathrm{lag}})$ versus $\log(m_{\mathrm{tot}})$, and obtains the exponent γ from the gradient). Explaining experimentally observed lag time versus concentration behavior is an important challenge for theoretical models, and different models make different predictions for the scaling exponent γ. In particular, if the lag time is dominated by the formation of a primary nucleus, we expect $\gamma > 1$, while for models where the lag time is dominated by autocatalytic growth by fragmentation, theory predicts that $\gamma = 1/2$ (8,11–13).

In this article, we show, using our own and others' experimental datasets, that lag time versus concentration curves are not always well-fitted by a simple power law. In contrast, we observe clear kinks in these curves, where the lag-time scaling changes. These kinks may reflect changes in the dominant molecular mechanism at play during the lag time. We explore using theoretical arguments and computer simulations whether the kinks can be explained by the interplay between the two mechanisms identified in existing models as being important during the lag phase: primary nucleation and autocatalytic growth. Our results show that the position and shape of the kinks are not well explained by the competition between nucleation and growth, as represented in existing models. This suggests that new mechanistic insights are needed to fully understand kinetic data on amyloid fibril assembly.

BACKGROUND

Theoretical models for amyloid fibril assembly

Theoretical models for the kinetics of amyloid fibril assembly often represent the key molecular processes as a set of chemical reactions. This reaction set typically includes primary nucleation, fibril elongation, and (possibly) fragmentation or secondary nucleation, and can be written as

$$n_c\, m \rightarrow f_{n_c}, \tag{1}$$

$$f_i + m \rightarrow f_{i+1}, \tag{2}$$

$$f_i \rightarrow f_j + f_{i-j}, \tag{3}$$

$$f_i + n_s\, m \rightarrow f_i + f_{n_s}. \tag{4}$$

Here, f_i denotes a fibril of length i, m denotes a protein monomer, and it is implicitly assumed that concentrations of all components are spatially homogeneous, i.e., that the system is well-mixed.

Reaction 1 describes the spontaneous formation of a primary nucleus (f_{n_c}), from n_c monomers (i.e., we assume that the nucleus consists of the smallest stable growth-competent fibril which is of length n_c). The total rate of this process is given by $(k_n/n_c!) \times m(t)^{n_c}$, where k_n is a nucleation rate constant and $m(t)$ is the time-dependent monomer concentration. This representation of nucleation as a single-step reaction whose rate depends on the n_cth power of m arises from a classical nucleation picture in which prenuclear aggregates are in rapid equilibrium before formation of the nucleus (17). The factor of $n_c!$ in the nucleation rate, which is not used in some other work (8), arises from ignoring the order in which the n_c monomers come together. This is equivalent to rescaling k_n.

Reaction 2 describes fibril growth by monomer addition at the ends. This process occurs at rate $2k_+m(t)f_i$, where k_+ is the elongation rate constant and f_i is the concentration of fibrils of length i. The factor of 2 arises because each fibril has two ends; omitting this amounts to a rescaling of k_+.

Reaction 3 describes fragmentation, in which a fibril of length i breaks into two shorter fibrils of length j and $i - j$. The fibril is assumed to break with equal probability at all sites along its length, so that the total rate of this process is given by $(i-1)k_f \times f_i$, where k_f is the fragmentation rate constant and the factor $(i = 1)$ arises because there are $i - 1$ possible breakage sites in a fibril of length i. In the case that one of the fragments is shorter than the nucleus size n_c, this fragment is assumed to split immediately into monomers—i.e., Reaction 3 should be replaced by $f_i \rightarrow j\, m + f_{i-j}$ or $f_i \rightarrow f_j + (i-j)\, m$, as appropriate.

Reaction 4 describes the autocatalytic formation of new fibrils by secondary nucleation on the surface of existing ones. Here, n_s monomers combine to form a fibril of length n_s, catalyzed by the presence of a fibril of length i. The rate at which this process happens is assumed to be $i(k_s/n_s!) \times m(t)^{n_s}$, where k_s is the secondary nucleation rate; this form of the rate equation is analogous to that for primary nucleation except for the factor of i, which accounts for the fact that longer fibrils have more sites where secondary nucleation can take place.

In this article, we base our theoretical analysis on the model described by Reactions 1–4 (using either the fragmentation reaction (3) or the secondary nucleation reaction (4), but not both). It is important to note, however, that others have also considered models including fibril end-joining (18), inhomogeneous fragmentation (19), and spatial propagation of the fibrillation process (20).

Predictions for lag-time scaling with protein concentration

The model described by Reactions 1–4 leads to a number of analytical predictions for the scaling of the lag time with protein concentration, in different parameter regimes. As in many experimental studies, we define the lag time as the time taken for the total amount of aggregated protein to reach a predetermined threshold (typically 10% of full aggregation).

Lag-time scaling without autocatalysis

We first discuss the case where new fibrils form only by primary nucleation (Reaction 1) and grow by elongation (Reaction 2)—i.e., the autocatalytic mechanisms, Reactions 3 and 4, are absent. This case has two possible regimes depending on the relative rates of primary nucleation and fibril elongation.

If primary nucleation is slow relative to the rate of fibril growth once a nucleus has formed, then we expect the lag time to be dominated by the time to formation of the first primary nucleus. As discussed above, the primary nucleation rate is assumed to be proportional to $m(t)^{n_c}$. At early stages in the aggregation process, almost all protein is present as monomers and $m(t) \approx m_{\mathrm{tot}}$. We therefore expect that the lag time is $\tau_{\mathrm{lag}} \sim m_{\mathrm{tot}}^{-n_c}$—i.e., in this regime the model predicts a lag-time versus concentration scaling exponent $\gamma = n_c$.

If, on the other hand, primary nucleation occurs on a similar or faster timescale than fibril growth, so that the fibril growth process makes a significant contribution to the lag time, a different prediction holds. In classic work, Oosawa and Asakura (21) and Oosawa and Kasai (22) showed that in this scenario, the kinetic curves for fibril growth can be described by

$$M(t) = m_{\mathrm{tot}}\left[1 - \mathrm{sech}^{2/n_c}\left[\left(\delta(n_c/2)^{1/2}\right)t\right]\right], \qquad (5)$$

where $M(t)$ is the total concentration of fibrillar protein and $\delta^{-1} \equiv (2k_+k_n m_{\mathrm{tot}}^{n_c}/n_c!)^{-1/2}$ defines a characteristic timescale. From Eq. 5, one can show that the lag-time scales with the protein concentration as $\tau_{\mathrm{lag}} \sim m_{\mathrm{tot}}^{-n_c/2}$—i.e., that in this regime the scaling exponent $\gamma = n_c/2$ (11,21). (Some versions of this model have instead $\gamma = (n_c + 1)/2$. This arises from a subtlety in the definition of the nucleus. Here we have assumed that the nucleus, of size n_c, is a growth-competent fibril. One could alternatively assume that the nucleus only becomes growth-competent upon addition of a further monomer. The latter scenario results in $\gamma = (n_c + 1)/2$.)

Lag-time scaling with autocatalysis

We now turn to the case where in addition to primary nucleation, and elongation, autocatalytic fibril formation can occur via fragmentation (Reaction 3) or secondary nucleation (Reaction 4). We assume that these mechanisms do not occur simultaneously—i.e., we have either Reaction 3 or Reaction 4 but not both. Knowles et al. (8) have presented an analytical solution to the deterministic differential equations corresponding to Reactions 1, 2, and either 3 or 4, under the assumptions that monomer depletion can be ignored at early times, and that $k_n << k_+$. This solution can be written as follows:

$$M(t) = m_{\mathrm{tot}}\left(1 - \exp\left(C_- e^{-\kappa t} - C_+ e^{\kappa t} + \frac{\delta^2}{\kappa^2}\right)\right). \qquad (6)$$

Here, two timescales come into play. The first is given by δ^{-1}, where δ is the same as in the Oosawa model (Eq. 5), and represents the characteristic timescale of fibril growth in the absence of autocatalysis. The second timescale is given by κ^{-1}, where κ takes different forms for fragmentation and secondary nucleation:

$$\kappa = \left(2k_+k_f m_{\mathrm{tot}}\right)^{1/2}$$

and

$$\kappa = \left(2k_+k_s m_{\mathrm{tot}}^{n_s+1}/n_s!\right)^{1/2},$$

respectively. This is the characteristic timescale for the autocatalytic growth process and depends on the rates of both growth and autocatalytic fibril formation. Finally, the constants $C_\pm$ are given by

$$C_\pm = (N_0k_+/\kappa) \pm \left[M_0/(2m_{\mathrm{tot}}) + \delta^2/\left(2\kappa^2\right)\right],$$

where N_0 and M_0 are the initial numbers of fibrils, and total fibrillar protein, per unit volume.

Defining a threshold concentration of aggregated protein M^*, we can compute from Eq. 6 the lag time τ_{lag} as the time at which M(t) reaches M^* (provided $M_0 << M^*$),

$$\tau_{\mathrm{lag}} = \frac{1}{\kappa}\log\left(\frac{D - \Phi + \left((\Phi - D)^2 + 4C_+C_-\right)^{1/2}}{2C_+}\right), \qquad (7)$$

where $D = \delta^2/\kappa^2$ and $\Phi = \log(1 - M^*/m_{\mathrm{tot}})$ (note that all logarithms are to base e). For unseeded aggregation, $M_0 = 0$ and $N_0 = 0$. If we further assume that the timescale of autocatalytic growth is faster than the Oosawa timescale, i.e., that $\delta << \kappa$, and that $M^* << m_{\mathrm{tot}}$, we obtain the simpler prediction

$$\tau_{\mathrm{lag}} \approx -\frac{1}{\kappa}\log\left[\frac{C_+}{\Phi^*_M}\right], \qquad (8)$$

where $\Phi^*_M = M^*/m_{\mathrm{tot}}$. In Eq. 8, the dominant contribution to the lag-time scaling comes from the dependence of the prefactor κ^{-1} on the protein concentration m_{tot} (while C_+ also varies with m_{tot}, the logarithm means that this contribution is weak). The value κ^{-1} scales differently with protein concentration for the two autocatalysis models. For fragmentation, $\kappa^{-1} \sim {m_{\mathrm{tot}}}^{-1/2}$, leading to a prediction for the lag-time scaling exponent $\gamma = 1/2$ (8,11–13). For secondary nucleation, $\kappa^{-1} \sim m_{\mathrm{tot}}^{-(n_s+1)/2}$, so that the model instead predicts $\gamma = (n_s + 1)/2$ (12).

MATERIALS AND METHODS

Experiments

Insulin sample preparation

Bovine insulin was obtained from Sigma-Aldrich, St. Louis, MO (I5,500, lot No. 0001434060) with zinc content ~0.5% (w/w). The samples in this

study were dissolved in 25 mM HCl (pH 1.6) immediately before the experiment. All solvents and solutions were filtered through a 0.22-mm filter (Millipore, Billerica, MA). Concentrations were checked via UV-vis absorption spectroscopy. ThT was added to each solution to a final concentration of 20 mM.

Lysozyme sample preparation

Lysozyme was obtained from Seikagaku, Tokyo, Japan (100940, lot No. E00 301), 6× crystallized. All samples in this study were dissolved in 20 mM HCl-KCl (pH2.0) buffer and 1 M NaCl, immediately before the experiment. All solvents and solutions were filtered through a 0.22 mm filter (Millipore). Concentrations were checked via UV-vis absorption spectroscopy. ThT was added to each solution to a final concentration of 55 μM.

ThT fluorescence kinetic measurements for bovine insulin and lysozyme

The ThT fluorescence measurements were conducted using a Fluostar plate reader (BMG Labtech, Life Technologies, Offenburg, Germany) and NBS 96-well plates (Cat. No. 3641; Corning, Corning, NY). Each well of the plate was filled with 100 μL of solution. Experiments for each protein concentration of bovine insulin were replicated across 2–3 whole plates, while those for lysozyme were replicated with several protein concentrations on the same plate. The final number of individual experiments for a given concentration of bovine insulin ranged from 140 to 200, and those of lysozyme ranged from 12 to 60. The plates were incubated at 60°C and fluorescence readings were taken from the bottom of the plate at wavelengths of 450 nm for excitation and 485 nm for emission. Readings were taken every 10 min in the case of bovine insulin, and 3 min for lysozyme.

Lag times for β_2-microglobulin

The lag-time data for β_2-microglobulin was kindly provided by Xue et al. (9). These data were obtained from ThT fluorescence kinetics at pH 2.0 and 37°C with 50 mM NaCl and 0.02% NaN_3. The data set contains 20 different protein concentrations in the range 8–244 μM, totaling 235 measurements.

Computer simulations

Simulation algorithm

We carried out stochastic computer simulations of the model defined by the chemical reaction sets 1–4 (including either Reaction 3 or Reaction 4, but not both), using a kinetic Monte Carlo algorithm (23). As we have shown in previous work, these simulations can quantitatively reproduce experimental kinetic curves (14). Here, we extend the parameter range, allowing us to explore the full range of possible behaviors of the model. Because our simulations are stochastic and account for discrete numbers of molecules, we can resolve individual molecular events such as the formation of the first primary nucleus. Because we focus on the lag time, simulations were terminated once half of the total protein was aggregated. For each parameter combination, 150 replicate simulation runs were performed.

Baseline parameter set

Our baseline parameter set for these simulations was obtained as in our previous work (14), by fitting Eq. 6 to averaged experimental data for the aggregation kinetics of bovine insulin at concentrations of 0.1, 0.2, 0.4, and 0.75 mg mL^{-1}, assuming fragmentation but not secondary nucleation, and a primary nucleus size $n_c = 2$. This fitting resulted in parameter values for fibril elongation and fragmentation $k_+ = 5 \times 10^4\ M^{-1}\ s^{-1}$ and $k_f = 3 \times 10^{-8}\ s^{-1}$. The primary nucleation rate k_n was varied in our simulations as described in Results. All protein was assumed to start in the monomeric form. For computational convenience, in most of our simulations we chose a small simulation volume $V = 0.83$ pL, comparable with that of a human cell.

Determination of lag times and scaling exponent

For each simulation trajectory, the total amount of fibrillar protein was obtained as a function of time by summing over all fibril lengths. The lag time was then determined as the time at which 3% of the total protein had been incorporated into fibrils (however, our results do not depend strongly on the choice of threshold). The lag time was computed for each simulation run and then averaged over replicate runs. To obtain the lag-time scaling exponent γ, we repeated our simulations over a range of protein concentrations m_{tot}, and using the Levenberg-Marquardt algorithm to fit our data to the functional form $\log(\tau_{lag}) = \log(A) - \gamma\log(m_{tot})$, where A is a constant, m_{tot} is the protein concentration and τ_{lag} denotes the lag time, averaged over replicate simulations, for a given protein concentration.

RESULTS

Experimental data shows kinks in lag time versus concentration plots

Fig. 1 shows lag time versus protein concentration data, obtained from our own experiments with bovine insulin and lysozyme (Fig. 1, *a* and *b*, respectively), and experiments by Xue et al. (9) on β_2 microglobulin Fig. 1 *c*). These experiments involved a large number of replicate experiments, and covered a wide range of protein concentrations (for details, see the caption of Fig. 1). These data show clear evidence for kinks: well-defined points in the lag-time versus concentration curves where the lag-time scaling changes. Because different lag-time scaling exponents are predicted to arise from different molecular mechanisms, this observation suggests that the dominant mechanism in early self-assembly kinetics changes with the protein concentration. Present-day theoretical models suggest that three mechanisms may be at play during the lag time: primary nucleation, sequential addition of monomers to existing fibrils, and autocatalytic growth via fragmentation or secondary nucleation. These mechanisms produce different lag-time scaling exponents (respectively, $\gamma = n_c$, $\gamma = n_c/2$, and $\gamma = 1/2$ or $(n_s + 1)/2$). Our starting hypothesis is therefore that the kinks that we observe in Fig. 1 arise from a crossover between regimes in which one or other of these mechanisms is dominant. In the rest of the article, we test this hypothesis, by assessing whether our models indeed predict kinks similar to those seen in the experimental data, first from a qualitative and then from a quantitative point of view.

Model predicts kinks due to competition between primary nucleation and autocatalysis

Simulations reveal changes in lag-time scaling exponent due to interplay between nucleation and growth

We first explore, for present-day theoretical models, how the interplay between different molecular mechanisms leads to changes in the lag-time scaling exponent. We carried out stochastic simulations of the chemical reaction sets Reactions 1–3, i.e., of a model which includes primary nucleation, fibril growth by elongation, and fragmentation (for

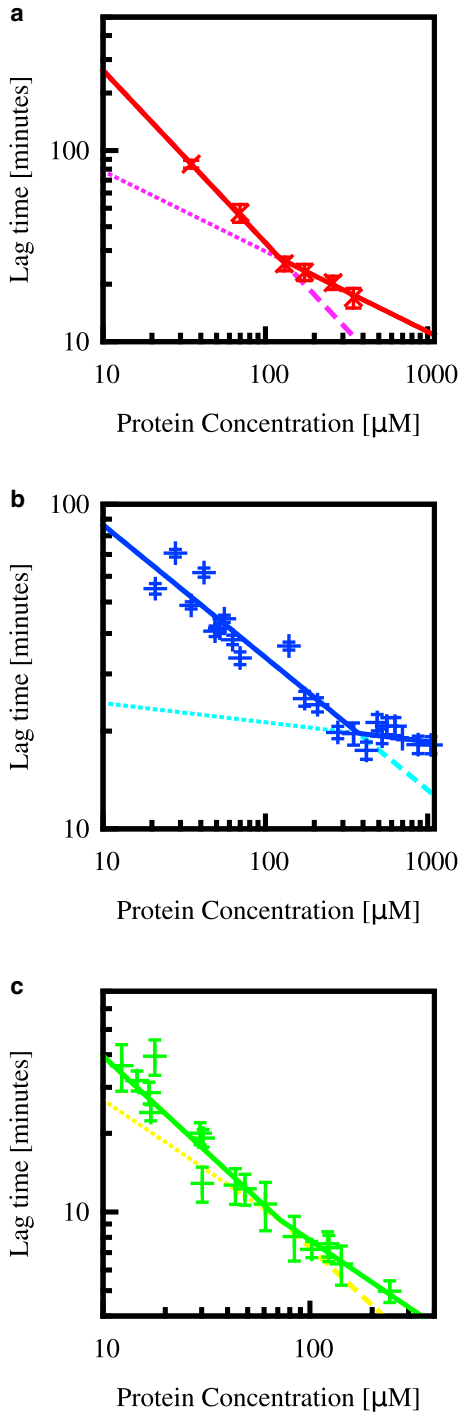

FIGURE 1 Lag time as a function of protein concentration for (*a*) bovine insulin, (*b*) lysozyme, and (*c*) β_2 microglobulin, in all cases averaged over replicate experiments. The data in panel *c* was kindly provided by Xue et al. (9). The error bars indicate the standard deviation among replicate experiments at a given concentration, and the lines indicate the best fit of the power law $\tau_{lag} \propto m_{tot}^{-\gamma}$ to data for a specified range of protein concentrations. (*a*) Two data sets for bovine insulin: (*black*) data fitted by the single power law over all protein concentrations (resulting in $\gamma = 0.36$) indicate experiments in the absence of NaCl; (*red*) data fitted by two power laws indicate self-assembly in the presence of 0.49 M NaCl. Each point corresponds to the mean and standard deviation of 140–200 kinetic traces at each of six protein concentrations (see Materials and Methods). Fitting the data for bovine insulin in the presence of NaCl in the range $m_{tot} \leq 160$ μM results in $\gamma = 0.90(2)$ (*dashed line*). Fitting the data in the range $m_{tot} \geq 160$ μM results in $\gamma = 0.42(10)$ (*dotted line*). (*b*) Data for lysozyme (for conditions, see Materials and Methods), where each point corresponds to the mean and standard deviation of 12–60 kinetic traces at each of 25 protein concentrations (see Materials and Methods). Fitting this data in the range $m_{tot} \leq 300$ μM results in $\gamma = 0.41(6)$ (*dashed line*); while fitting the data in the range $m_{tot} \geq 300$ μM results in $\gamma = 0.06(5)$ (*dotted line*). (*c*) Data for β_2 microglobulin, where each point corresponds to the mean and standard deviation of 235 kinetic traces at 20 different protein concentrations (for conditions, see Materials and Methods and Xue et al. (9)). Fitting the data in the range $m_{tot} \geq 30$ μM gives $\gamma = 0.77(14)$ (*dashed line*), while fitting the data in the range $m_{tot} \geq 40$ μM gives $\gamma = 0.54(5)$ (*dotted line*). To see this figure in color, go online.

simulation details, see Methods). For now we neglect secondary nucleation, but this will be considered later. As a simple way to control the interplay between different molecular mechanisms, we systematically varied the primary nucleation rate constant k_n over a wide range (15 orders of magnitude), keeping all other parameters fixed. Primary nucleation rates are poorly determined by fitting theoretical predictions such as Eq. 6 to experimental data (8), so reliable estimates of their true values are lacking, but the information that is available does suggest a wide range of possible nucleation rate constants: 10^{-4}–10^{-14} s^{-1} (8,24,25) for various protein systems. For each value of the nucleation rate constant k_n, we repeated our simulations for protein concentrations m_{tot} in the range $1 \times 10^{-5} \leq m_{tot} \leq 7.5 \times 10^{-4}$ M (commensurate with concentration ranges used in our own and others' experiments (9,14)), and computed the lag-time scaling exponent γ using linear fits to our simulation data for $\log(\tau_{lag})$ versus $\log(m_{tot})$, as described in Materials and Methods.

Fig. 2 *a* shows plots of lag-time versus concentration obtained in our simulations, for increasing values of the primary nucleation rate constant (top to bottom; also color-coded *purple* to *red*). As expected, we see clear changes in lag-time scaling behavior as the nucleation rate constant changes, reflecting the shift in relative importance of primary nucleation and growth in determining the lag time. These changes in lag-time scaling can be seen in more detail in Fig. 2 *b*. Here, the black symbols (with *error bars*) show the scaling exponent γ, extracted from the gradients of the curves in Fig. 2 *a*, plotted as a function of the primary nucleation rate constant k_n (note that, for clarity, Fig. 2 *a* shows only a subset of our simulated nucleation rates). Two features are immediately apparent.

1) The limits of high and low nucleation rate constants are consistent with theoretical predictions. If nucleation is very slow, γ tends to 2; this is consistent with the prediction $\gamma = n_c$ (= 2 in this case), for the regime in which the lag time is dominated by the time to formation of the first primary nucleus. In contrast, if nucleation is very rapid, we obtain $\gamma = 1$, consistent with the Oosawa prediction $\gamma = n_c/2$ of nucleation-dependent polymerization in the absence of autocatalysis.

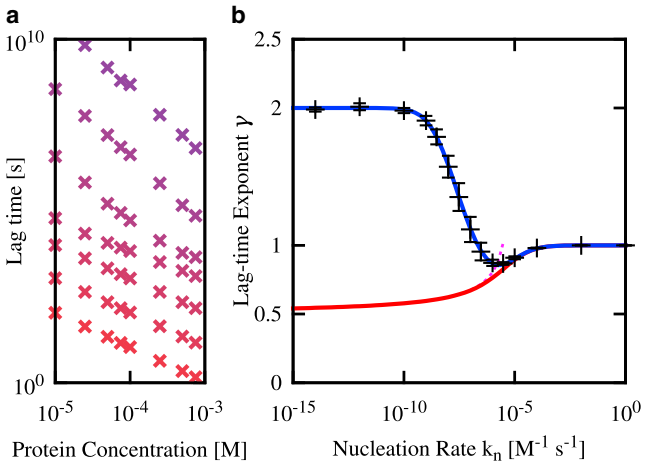

FIGURE 2 Lag-time scaling with protein concentration for the fragmentation model depends on the primary nucleation rate constant. (*a*) Lag-time versus concentration curves obtained from stochastic simulations, for several values of the nucleation rate constant k_n: k_n increases top to bottom in factors of 100, from 10^{-12} M^{-1} s^{-1} (*purple*) to 1 M^{-1} s^{-1} (*red*). (*b*) Lag-time scaling exponent γ as a function of primary nucleation rate constant k_n. (*Black symbols*) Simulation results obtained by fitting the data shown in panel *a* to $\log(\tau_{lag}) = \log(A) - \gamma\log(m_{tot})$. The error bars are dominated by the confidence in the fit, rather than the error in the lag-time measurements. (*Red line*) The value $\gamma(k_n)$ extracted from the full prediction of the fragmentation model, Eq. 7. (*Magenta line*) The value $\gamma(k_n)$ extracted from the approximate prediction of the fragmentation model, Eq. 8. (*Blue line*) The value $\gamma(k_n)$ extracted from the lag-time expression Eq. 9—i.e., the full fragmentation model solution augmented by an extra term to describe the first primary nucleation event. For all the data shown, the protein concentration range was $1 \times 10^{-5} \le m_{tot} \le 7.5 \times 10^{-4}$ M and the parameters were $n_c = 2$, $k_+ = 5 \times 10^4$ M^{-1} s^{-1}, $k_f = 3 \times 10^{-8}$ s^{-1}, and the volume $V = 0.83$ pL. To see this figure in color, go online.

2) For intermediate values of the nucleation rate constant, the scaling exponent dips down toward the prediction of the fragmentation model, $\gamma = 1/2$, but never actually reaches this value.

These results show clearly that changes in the dominant molecular mechanism can indeed produce shifts in the scaling exponent. Interestingly, they also show that the regime where autocatalysis (in the form of fibril fragmentation) is dominant can become masked by the effects of primary nucleation, leading to an apparent scaling exponent significantly larger than one-half, even for systems where fragmentation plays an important role.

Modification of standard theory is needed to reproduce simulation results

Fig. 2 *b* (*red line*) also shows the lag-time scaling exponents emerging from the analytical expression for the lag time, Eq. 7, derived for a model that includes nucleation, fibril elongation, and fragmentation, all treated deterministically. Lag-times derived from the limit of this expression in the case of rapid autocatalysis (Eq. 7; *magenta line*) produce similar results. These results were obtained by using the analytical expressions to plot $\log(\tau_{lag})$ versus $\log(m_{tot})$, from which predictions for γ were extracted numerically, for different values of the nucleation rate constant k_n. The predictions of this deterministic model are in good agreement with our simulation results for higher nucleation rate constants, showing that the analytical solution to the fragmentation model does correctly capture the crossover to an Oosawa-like regime. (The crossover to the Oosawa-like regime can be seen in the analytical expression Eq. 7 by noting that at high values of k_n, $\delta >> \kappa$, and in this limit D in Eq. 7 becomes large, which gives $\tau_{lag} \approx (2\Phi)^{1/2}\,\delta^{-1}$ and $\gamma = 1$.) However, the deterministic model does not capture the shift to the primary nucleation-dominated regime ($\gamma = n_c = 2$) at low values of k_n, instead tending toward the fragmentation-like value of $\gamma = 1/2$ as k_n decreases.

Why does the analytical result fail to capture the simulation behavior at low nucleation rates? This discrepancy arises because the deterministic differential equations used to generate the theoretical prediction, Eq. 7, do not take account of the discrete nature of the primary nucleation process. In the deterministic differential equation representation, all concentrations are continuous variables that can be arbitrarily small—thus the concentration of nuclei increases continuously from time zero and initially corresponds to less than one nucleus in the corresponding simulated volume. In the continuous model, the autocatalytic growth processes are able to operate on this very low concentration of nuclei, leading to a lag time that is dominated by κ (the timescale of autocatalytic growth) rather than by slow primary nucleation. In contrast, in our stochastic simulations, as in reality, the autocatalytic growth processes cannot start until the first nucleus has been generated—thus the lag time is dominated by the formation of the first primary nucleus, leading to $\gamma = n_c = 2$.

Fortunately, it turns out that this problem can easily be remedied. The formation of the first primary nucleus in the volume V occurs as a Poisson process, with mean waiting time $n_c!(VN_Ak_nm_{tot}^{n_c})^{-1}$, where V is the volume and N_A is Avogadro's number (this follows from our definition of the nucleation reaction, Reaction 1). In the regime where the nucleation rate is very low, we expect that a single nucleus forms, and grows autocatalytically until the threshold is reached (i.e., we do not expect to see multiple primary nucleation events). Once the nucleus has formed, its growth should be well described by the deterministic model prediction, Eq. 7. We can therefore simply add the predicted waiting time for formation of a single nucleus to Eq. 7 to obtain the new prediction

$$\tau_{lag} = \frac{1}{\kappa}\log\left(\frac{D - \Phi + \left((\Phi - D)^2 + 4C_+C_-\right)^{1/2}}{2C_+}\right) + \frac{n_c!}{VN_Ak_nm_{tot}^{n_c}},$$

where, as before, $D = \delta^2/\kappa^2$ and $\Phi = \log(1 - M^*/m_{tot})$. Because we are considering the unseeded case,

$C_{\pm} = \pm\delta^2/\kappa^2 = \pm D/2$, we can rewrite Eq. 7. Modification of standard theory is needed to reproduce simulation results, thus:

$$\tau_{\text{lag}} = \frac{1}{\kappa}\log\left(1 - \Phi D^{-1} + \left(\left(\Phi D^{-1}\right)^2 - 2\Phi D^{-1}\right)^{\frac{1}{2}}\right) + \frac{n_c!}{VN_A k_n m_{\text{tot}}^{n_c}}.$$

If k_n is large such that $\delta >> \kappa$, we expect the second term in Eq. 7. Modification of standard theory is needed to reproduce simulation results to become negligible, leading to the same scaling behavior as Eq. 8 (*red line* in Fig. 2 *b*).

If, on the other hand, k_n is small such that $\delta << \kappa$, Eq. 9 can be simplified by neglecting small terms in the logarithm:

$$\tau_{\text{lag}} \approx \frac{1}{\delta^2}\left[D\kappa\log\left(\Phi D^{-1}\right) + \frac{2k_+}{VN_A}\right]. \quad (9)$$

As $k_n \rightarrow 0$, the logarithmic term can be neglected completely, yielding $\tau_{\text{lag}} \approx 2k_+/(VN_A\delta^2)$ and hence $\gamma = n_c$ as expected when primary nucleation is very slow. For large nucleation rates k_n, the logarithmic term dominates, leading to the autocatalysis-dominated scaling regime with $\gamma = 1/2$.

The blue line in Fig. 2 *b* shows the lag-time scaling exponent extracted from the modified lag-time expression (9), as a function of the nucleation rate constant k_n. This prediction is indeed in excellent agreement with our simulation results over the entire range of primary nucleation rates.

Theory predicts kinks in lag-time versus concentration plots

Using the modified analytical lag-time expression, Eq. 9, which correctly accounts for the stochasticity of primary nucleation, we now investigate whether present-day theoretical models can account for the kinks that we observe in experimental lag time versus concentration plots. Fig. 3 *a* shows lag-time versus concentration curves predicted by Eq. 9, plotted over a wide range of protein concentrations (nine orders of magnitude). Results are shown for various values of the nucleation rate constant k_n, indicated by the colors; k_n increases top-right to bottom-left. The symbols show our stochastic simulation results, which are in good agreement with the predictions of Eq. 9. It is immediately apparent from Fig. 3 *a* that the model does indeed predict kinks in the lag-time versus concentration curves. These kinks appear for intermediate values of the nucleation rate constant. In these curves, the lag-time scaling changes smoothly, over approximately an order of magnitude in protein concentration, between the fragmentation-dominated scaling $\gamma = 1/2$ and either the primary nucleation-dominated scaling $\gamma = n_c$ or the Oosawa (elongation-dominated) scaling $\gamma = n_c/2$. For very low or very high values of the nucleation rate constant, the lag-time versus concentration

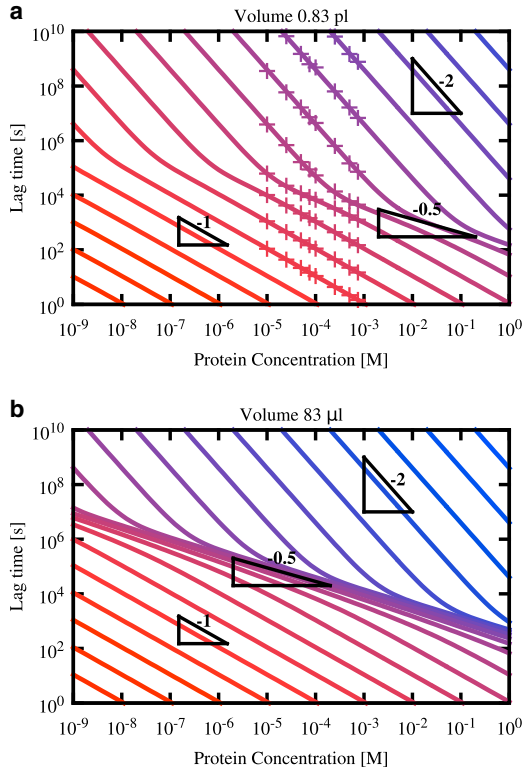

FIGURE 3 Transitions between different scaling regimes appear as kinks in lag-time versus concentration curves. Lag-time versus concentration curves predicted by Eq. 9 are plotted for a range of primary nucleation rates k_n. The k_n value increases top-right to bottom-left and is color-coded from $k_n = 10^{-30}$ M^{-1} s^{-1} (*blue*) to $k_n = 10^{10}$ M^{-1} s^{-1} (*red*), being increased by factors of 100. The two panels show results for different system volume V: (*a*) $V = 0.83$ pL, (approximately the size of a human cell). (*b*) The much larger volume $V = 83$ μL (a typical volume used in in vitro protein aggregation experiments). In panel *a*, the crosses correspond to the simulation data given in Fig. 2 *a*. In all cases, $k_+ = 5 \times 10^4$ M^{-1} s^{-1}, $k_f = 3 \times 10^{-8}$ s^{-1}, and $n_c = 2$. To see this figure in color, go online.

curves in Fig. 3 *a* do not show kinks, showing instead either primary nucleation-dominated scaling, $\gamma = n_c = 2$, or Oosawa scaling, $\gamma = n_c/2 = 1$, respectively. Thus our analysis shows that present-day models are consistent with the occurrence of kinks in lag-time versus concentration curves. These curves may show no kinks, one kink, or two kinks, depending on the molecular parameters and the protein concentration range observed.

Fragmentation-dominated scaling is more apparent at larger volumes

In vitro protein aggregation experiments are typically performed using system volumes of between 10 μL and 10 mL (10^{-5} to 10^{-2} L), while the typical volume of a human cell is $\leq$10–100 pL (10^{-11}–10^{-10} L). To investigate the effect of sample volume, we repeated our analysis for a larger volume of 83 μL. For this volume, stochastic simulations are not practical, but the analytical result of Eq. 9 still holds. We expect that changing the sample volume

will have a strong effect on the kinks in the lag-time versus concentration curves, because the time to formation of the first primary nucleus scales inversely with the volume (see Eq. 9). However, once a nucleus has been formed, the average time for the autocatalytic growth process to reach the threshold is independent of the volume (see, for example, Eq. 7). This implies that changing the system volume will shift the crossover points between the regimes in which the lag time is dominated by primary nucleation and by fragmentation—i.e., it will shift the kinks in the lag-time versus concentration curves. For larger systems, we expect the fragmentation-dominated regime to extend over a wider range of protein concentrations than for smaller systems.

Fig. 3 *b* shows lag-time versus concentration curves, computed using Eq. 9 for a volume of 83 μL, for the same parameter set as in Fig. 3 *a*. For the larger volume, we still see the two primary nucleation-dominated regimes $\gamma = n_c$ and $\gamma = n_c/2$ at low and high nucleation rates, respectively. However, for intermediate nucleation rates, the fragmentation-dominated scaling regime $\gamma = 1/2$ typically extends over a wide range of concentrations (for some nucleation rates, we even see $\gamma = 1/2$ over the entire nine orders of magnitude of protein concentration).

Interestingly, these results suggest that lag-times for amyloid fibril formation are likely to scale quite differently with protein concentration in small volume samples than in in vitro experiments where the volume is large. While fragmentation-dominated lag-time scaling, with exponent $\gamma = 1/2$, may be a common feature of in vitro experiments, it is likely to be much less prevalent in smaller volumes on the scale of human cell, which are of clinical relevance.

Similar results are obtained for autocatalysis via secondary nucleation

Up to now, we have focused on the case where the autocatalytic growth mechanism is provided by fibril fragmentation. We now show that similar phenomena occur when autocatalysis instead occurs by secondary nucleation on the surface of existing fibrils. Fig. 4 *a* shows predicted lag-time versus concentration curves, for different values of the nucleation rate constant, for a model with primary nucleation, fibril elongation, and secondary nucleation (i.e., Reactions 1, 2, and 4). In this case, the same theoretical prediction for the lag time, Eq. 9, holds, but we use the alternative expression for the timescale of autocatalytic growth, $\kappa = (2k_+k_s m_{\mathrm{tot}}^{n_s+1}/n_s!)^{1/2}$. The predicted values of γ in the primary nucleation-dominated regimes are the same as for the fragmentation case: for very low nucleation rates (or low protein concentrations) we expect $\gamma = n_c = 2$ while for high nucleation rates (or high protein concentrations) we expect $\gamma = n_c/2 = 1$. However, the scaling in the intermediate, autocatalysis-dominated regime is now predicted to be $\gamma = (n_s + 1)/2$, where n_s is the size of the secondary nucleus (here assumed to be $n_s = 2$ so that we predict

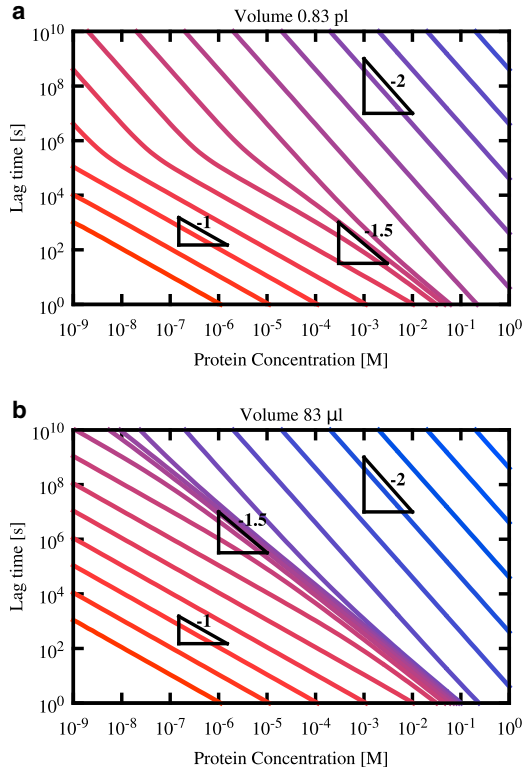

FIGURE 4 Kinks are also predicted in lag-time versus concentration curves for a model where the autocatalysis mechanism is provided by secondary nucleation. Lag-time versus concentration curves predicted by Eq. 9 are shown for a series of values of the primary nucleation rate k_n (increasing *top-right* to *bottom-left*) from 10^{-30} M^{-1} s^{-1} (*blue*) to 10^{6} M^{-1} s^{-1} (*red*), increasing by factors of 100. The two panels show results for different system volume V: (*a*) $V = 0.83$ pL (approximately the size of a human cell). (*b*) The much larger volume $V = 83$ μL (a typical volume used in in vitro protein aggregation experiments). The other parameters are $k_+ = 5 \times 10^4$ M^{-1} s^{-1}, $k_s = 24$ M^{-1} s^{-1}, $n_c = 2$, and $n_s = 2$. To see this figure in color, go online.

$\gamma = 3/2$). These three regimes are indeed apparent in Fig. 4 *a*. As in the fragmentation case, increasing the system size increases the range of parameters (i.e., k_n and protein concentration) for which autocatalysis dominates the lag time (Fig. 4 *b*).

Interestingly, the predicted lag-time scaling in the high concentration limit is somewhat different in the case of secondary nucleation, compared to fragmentation. In the fragmentation case, at very high protein concentrations, we expect to obtain the Oosawa-like scaling $\gamma = n_c/2$ (see Fig. 3). In contrast, for secondary nucleation, some of our lag time versus concentration curves never reach the Oosawa-like scaling limit but instead converge to the secondary nucleation-dominated scaling exponent $\gamma = (n_s + 1)/2 = 3/2$ in the high concentration limit (this is most apparent in Fig. 4 *b*). The crossover between the Oosawa-like and autocatalysis-dominated regimes occurs when $\delta = \kappa$. These two parameters scale differently with protein concentration: δ scales as $m_{\mathrm{tot}}^{n_c/2}$ while κ scales as

$m_{\mathrm{tot}}^{(n_s+1)/2}$ for secondary nucleation and $m_{\mathrm{tot}}{}^{1/2}$ for fragmentation. For fragmentation, because $n_c/2 > 1/2$, at high enough protein concentration there will always be a regime where $\delta > \kappa$, leading to Oosawa-like scaling. For secondary nucleation, however, we may be in a situation where $n_c/2 < (n_s + 1)/2$ (this is indeed the case for our parameter set). In this case, for some parameter combinations, there may be no protein concentration range for which $\delta > \kappa$; thus the Oosawa-like scaling regime may never be reached. In this case, we expect to see only a single kink in the lag time versus concentration curve, at the transition between the primary nucleation- and secondary nucleation-dominated regimes.

Are the kinks in experimental data explained by contemporary models?

We now assess quantitatively whether the kinks that we observe in the experimental data of Fig. 1 are consistent with the shifts among regimes dominated by primary nucleation, autocatalytic growth, and sequential monomer addition, as predicted by the model.

We first consider the values of the scaling exponent to the left and right of the kink, which we denote γ_L and γ_R. For insulin (Fig. 1 *a*), we obtain $\gamma_L = 0.90(2)$ and $\gamma_R = 0.42(10)$. These exponents are roughly consistent with a shift from a primary nucleation-dominated regime at low concentration, with $\gamma_L = n_c \approx 1$ to a fragmentation-dominated regime at higher concentration, with $\gamma_R \approx 1/2$. Similarly, the exponents for β_2 microglobulin (Fig. 1 *c*), $\gamma_L = 0.77(14)$ and $\gamma_R = 0.54(5)$, are roughly consistent with the same scenario. In contrast, for lysozyme (Fig. 1 *b*), we obtain quite different exponents, $\gamma_L = 0.41(6)$ and $\gamma_R = 0.06(5)$. In this case, while γ_L could conceivably correspond to a fragmentation-dominated regime ($\gamma \approx 1/2$), the observed value of $\gamma_R = 0.06(5)$ appears to be too small to correspond to any of the regimes predicted by the model. We also note that in the case where $n_c = 1$, our model predicts that the Oosawa regime of nucleation-dependent polymerization will have a lag-time scaling exponent $\gamma = 1/2$ and thus not be distinguishable from the fragmentation-dominated regime. However, the shape of the growth curves at protein concentrations to the right of the kink (9,14) still appear strongly dominated by autocatalysis, and so it is unlikely that our range of data encompasses the second kink, i.e., a transition to the Oosawa regime.

We next consider the protein concentration at which the kink happens (the position of the kink) in the plots of Fig. 1. If the kinks do indeed signify a transition between primary nucleation- and fragmentation-dominated regimes (at least for insulin and β_2 microglobulin), we can use our analytical model results to predict the concentration at which this should happen. Returning to our lag-time prediction, Eq. 9, the two terms inside the square brackets arise from autocatalytic growth and primary nucleation, respectively. The crossover in dominance between these mechanisms should happen when these two terms are equal in magnitude, i.e., when the following condition is satisfied:

$$\frac{2k_+}{V\,N_A} = D\kappa\log\left(\Phi^{-1}D\right). \tag{10}$$

Here, $\Phi = \log(1 - M^*/m_{\mathrm{tot}})$ is a constant relating to the lag-time threshold M^*, and $D = \delta^2/\kappa^2$ measures the relative importance of the characteristic timescales for Oosawa-like growth $\delta^{-1} \equiv (2k_+k_n m_{\mathrm{tot}}^{n_c}/n_c!)^{-1/2}$ and fragmentation-dominated growth $\kappa^{-1} = (2k_+k_f m_{\mathrm{tot}})^{-1/2}$. We now would like to determine whether Eq. 10 is satisfied for the kinks that we observe in the experimental data of Fig. 1.

A direct test of Eq. 10 requires knowledge of the kinetic parameters κ, k_+, and k_n, for each of the protein systems shown in Fig. 1. The value of κ can be extracted to a high degree of accuracy by fitting the predictions of the deterministic model, Eq. 6, to kinetic curves, and it can be reasonably assumed that the elongation rate k_+ takes a similar value to that of insulin, for which it is known (14). Kinetic fits do not, however, accurately determine the nucleation rate constant k_n, which is believed to vary widely among proteins (8,24,25). We are therefore obliged to take an indirect approach to testing Eq. 10. We ask what value of the nucleation rate constant k_n would be required in order for Eq. 10 to be satisfied, for the data of Fig. 1. Then we ask whether this resulting prediction for k_n is reasonable. If it is not reasonable, we can conclude that the kinks observed in the data of Fig. 1 cannot be explained by a transition between primary nucleation- and autocatalytic growth-dominated lag time regimes.

To this end, we extracted from the datasets of Fig. 1 the protein concentration (i.e., the value of m_{tot}) at which the kink was observed. Using these values together with the known values of the sample volume V and lag-time threshold M^*, values of κ obtained by fitting kinetic curves (either in our own work (14) or that of Xue et al. (9)), and the value of k_+ obtained by fitting the kinetic curves of insulin (14), we solved Eq. 10 numerically for D. From this, we obtained a value of the nucleation rate constant consistent with the position of the kink on the protein concentration axis, for each dataset. Based on our observations for the scaling exponents left of the kink, γ_L, we chose to use a nucleus size $n_c = 1$, but we also repeated our calculations for larger values of n_c. The parameters used in these calculations, together with the resulting values of k_n, are presented in Table 1. Note that the predicted value of k_n has different dimensionality depending on the assumed nucleus size.

To determine whether the resulting predictions for the nucleation rate constant k_n were reasonable, we used them to calculate an estimated time to formation of the first nucleus, in a sample of the same volume and protein concentration as the experiment. This time is given by $n_c!/[k_n m_{\mathrm{tot}}^{n_c} N_A V]$, and, due to a cancellation with the n_c-dependence of k_n, it is independent of the chosen value of n_c. Table 1 lists the resulting times to formation of the first

TABLE 1 Testing whether the condition from Eq. 10 is consistent with the kinks observed in the experimental data of Fig. 1

Description	Insulin	Lysozyme	β_2 microglobulin
m_{tot} at kink (from Fig. 1)	160 μM	300 μM	40 μM
κ from kinetic fit (9,14)	1.3×10^{-4} s^{-1}	1.2×10^{-3} s^{-1}	7.2×10^{-5} s^{-1}
k_+ from insulin kinetic fit (14)	5×10^{4} M^{-1} s^{-1}	5×10^{4} M^{-1} s^{-1}	5×10^{4} M^{-1} s^{-1}
V	100 μL	100 μL	100 μL
M^*/m_{tot}	0.1	0.1	0.1
Solution for k_n for $n_c = 1$	5.2×10^{-22} M^{-1} s^{-1}	2.3×10^{-21} s^{-1}	1.2×10^{-21} s^{-1}
Solution for k_n for $n_c = 2$	6.4×10^{-18} M^{-1} s^{-1}	1.6×10^{-17} M^{-1} s^{-1}	5.8×10^{-17} M^{-1} s^{-1}
Solution for k_n for $n_c = 3$	1.2×10^{-13} M^{-1} s^{-1}	1.6×10^{-13} M^{-1} s^{-1}	4.4×10^{-12} M^{-1} s^{-1}
Average time to nucleus $= n_c!/[k_n m_{tot}^{n_c} N_A V]$	56 h	7 h	98 h
Predicted lag time at kink	112 h	13 h	197 h
Lag time at kink (from Fig. 1)	~25 h	~20 h	~10 h

rows 2–6 show the parameter values used. Rows 7–9 show the resulting values of the nucleation rate constant k_n from numerical solution of Eq. 10, for $n_c = 1$, $n_c = 2$, and $n_c = 3$. Rows 10–12 show the implied predictions for the time to formation of the first nucleus and the lag time at the kink. The latter is compared to the experimental value extracted from Fig. 1. Values of κ were extracted from the gradient of the steepest part of the growth curves. Those shown here are the means, for each protein, at the protein concentration where the kink occurs.

nucleus. For insulin and β_2 microglobulin, it is immediately apparent that the predicted waiting times are not reasonable, because they are significantly longer than the observed lag time at the kink. We can also directly calculate the predicted lag time at the kink, by inputting our predicted value of k_n (together with the other parameters) into Eq. 9. The results are presented in Table 1 and compared with the observed lag times at the kink, extracted from the data of Fig. 1 (again, these predictions are independent of n_c).

Again, for insulin and β_2 microglobulin, the predicted lag time is far longer than the observed lag time, suggesting that Eq. 10 is not satisfied for a reasonable choice of primary nucleation rate constant. Interestingly, for lysozyme, this procedure actually predicts a lag time that is within a factor of 2 of that observed experimentally (and is shorter rather than longer than the experimental value). While this might suggest that the kink observed for lysozyme could be explained by a transition between primary nucleation- and autocatalytic growth-dominated regimes, we note that for lysozyme, unlike the other proteins, the scaling exponents to the left and right of the kink do not seem to be consistent with such a transition.

To conclude, our analysis shows that the kinks we observe in the experimental lag-time versus concentration plots of Fig. 1 are not quantitatively consistent with a transition from primary nucleation- to growth-dominated lag-time regimes. For bovine insulin and β_2 microglobulin, such a transition requires a primary nucleation rate constant that is too small to be consistent with the data, while for lysozyme the scaling exponents to the left and right of the kink are inconsistent with this transition. Understanding the origin of these kinks is likely to require new mechanistic understanding, beyond that provided by existing kinetic models.

DISCUSSION

In experimental studies of amyloid fibril self-assembly kinetics, measurements of lag time as a function of protein concentration are often used to diagnose the underlying molecular mechanism (26). Based on predictions of theoretical models that include primary nucleation, elongation by monomer addition, and autocatalytic growth via fragmentation (or secondary nucleation), a scaling exponent $\gamma < 1$ is usually taken to imply a fragmentation-dominated mechanism, while an exponent $\gamma > 1$ suggests that fragmentation is not involved. In this article, we find that a more complex picture emerges from experimental lag-time versus concentration curves, when they are averaged over many replicates and measured over a wide range of protein concentrations. For three different protein systems (including both our own and others' data), we observe kinks in the lag-time versus concentration curves, at which the scaling exponent γ changes. Other published datasets also show tantalizing hints of kinks, but with insufficient statistical certainty, or without a sufficiently wide concentration range to be sure (9,27–29).

The presence of a kink in the lag-time versus concentration curve apparently signifies a change in the dominant molecular mechanism at play during the lag phase of fibril assembly. To test whether these kinks are consistent with the existing theoretical picture, we carried out a detailed analysis of the lag-time scaling behavior of the standard model, including nucleation, elongation, and fragmentation (or secondary nucleation). Our analysis shows that indeed this model does predict changes in lag-time versus concentration scaling, and that these kinks in the lag-time versus concentration curve are due to crossovers among primary nucleation-dominated, autocatalysis-dominated, and elongation-dominated regimes. Capturing these kinks correctly requires us to modify the deterministic expression for the lag time to take account of stochastic primary nucleation events.

Importantly, however, an inconsistency is revealed upon quantitative comparison between the properties of the kinks that we observe in our and others' experimental data, and the model predictions. For our data on bovine insulin (14)

(Fig. 1 *a*) and the data of Xue et al. (9) on β_2 microglobulin (Fig. 1 *c*), the scaling exponents to the left and right of the kink are broadly consistent with a transition from a primary nucleation-dominated regime at low protein concentration to a fragmentation-dominated regime at higher concentration. However, fitting the protein concentration at which this transition is observed to the prediction of the theoretical model requires us to assume a primary nucleation rate constant that is too low to account for the observed magnitude of the lag time. Thus for these protein systems, the theoretical model prediction is not quantitatively consistent with the data. For lysozyme (Fig. 1 *b*) the model also fails to account for the data, but in a different way: while the position of the kink is predicted correctly within a factor of 2, the scaling exponent to the right of the kink is inconsistent with any of the regimes predicted by the model. Thus, while the standard theoretical model generally does a good job of explaining kinetic curves for amyloid fibril formation, new mechanistic understanding is needed to explain detailed features of the lag-time concentration curves.

What mechanisms could account for the kinks that we see in the data of Fig. 1? The standard theoretical model investigated in this work is incomplete in that it does not include fibril end-joining (18) or length-dependent fragmentation; however, neither of these factors is expected to have a significant influence during the lag phase (30). Another factor not included in our model is spatial heterogeneity. For insulin fibril formation tracked microscopically in microdroplets, fibril formation has been shown to propagate out from an initial nucleation site as a spatial wave (20,31). It would be very interesting to investigate the consequences of this spatial propagation for the apparent lag-time scaling; however, it seems unlikely that it would affect the scaling in the apparent nucleation-dominated regime (left of the kink), where we obtain lower-than-expected scaling exponents for insulin and β_2 microglobulin. It seems possible that our results may reflect the nature of the primary nucleation event itself. While this is often assumed to be a homogeneous, stochastic event occurring at a rate proportional to m^{n_c} (i.e., involving the spontaneous union of n_c monomers), fitting in vitro kinetic data is actually highly sensitive to the nature of the nucleation event. Factors such as nucleation on the surface of the sample chamber, the presence of preexisting nuclei, or slow conformational changes during the nucleation process could result in lag-time scaling with exponents lower than those predicted by the standard model. More generally, high-concentration saturation, either of the primary nucleation step (e.g., due to surface nucleation) or of the fibril elongation process, seems likely to lead to decreased lag-time scaling exponents. Indeed, lag-time scaling exponents smaller than one-half are quite frequently observed both in our own experimental work (14), and in that of others (8,16,32–36).

From an immediate practical point of view, our results demonstrate that great care is needed in interpreting lag-time scaling exponents from protein assembly data. More broadly, our work highlights a need to better understand the molecular mechanisms at play during the lag phase of amyloid fibril assembly. The lag phase of fibril assembly is of particular importance in the clinical context, both because lag-time variability may be associated with variable time of neurodegenerative disease onset (30) and because it has been suggested that toxicity of early-stage aggregates rather than fully assembled fibrils may be causative in disease (5,37,38). Detailed measurements of lag-time versus concentration scaling, combined with the development of new mechanistic models, may provide a way to probe what is going on during this crucial stage of the assembly process.

ACKNOWLEDGMENTS

We thank Juraj Szavits-Nossan for useful discussions. The β_2 microglobulin data was used in our analysis with kind permission of Xue et al. (9).

This work was supported by the Engineering and Physical Sciences Research Council under grant No. EP/J007404. K.E. was supported by an Engineering and Physical Sciences Research Council Doctoral Training Partnership endowment and R.J.A. was supported by a Royal Society University Research Fellowship.

REFERENCES

1. Chiti, F., and C. M. Dobson. 2006. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75:333–366.
2. Knowles, T. P. J., and M. J. Buehler. 2011. Nanomechanics of functional and pathological amyloid materials. *Nat. Nanotechnol.* 6: 469–479.
3. Morris, R. J., and C. E. MacPhee. 2013. Amyloid protein biomaterials. *In* Encyclopedia of Biophysics. Springer, New York, pp. 76–81.
4. Kayed, R., E. Head, …, C. G. Glabe. 2003. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science.* 300:486–489.
5. Glabe, C. G., and R. Kayed. 2006. Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis. *Neurology.* 66:S74–S78.
6. Glabe, C. G. 2008. Structural classification of toxic amyloid oligomers. *J. Biol. Chem.* 283:29639–29643.
7. Binger, K. J., C. L. L. Pham, …, G. J. Howlett. 2008. Apolipoprotein C-II amyloid fibrils assemble via a reversible pathway that includes fibril breaking and rejoining. *J. Mol. Biol.* 376:1116–1129.
8. Knowles, T. P. J., C. A. Waudby, …, C. M. Dobson. 2009. An analytical solution to the kinetics of breakable filament assembly. *Science.* 326:1533–1537.
9. Xue, W.-F., S. W. Homans, and S. E. Radford. 2008. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc. Natl. Acad. Sci. USA.* 105:8926–8931.
10. Ferrone, F. A. 1999. Analysis of protein aggregation kinetics. *Methods Enzymol.* 309:256–274.
11. Cohen, S. I. A., M. Vendruscolo, …, T. P. J. Knowles. 2011. Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments. *J. Chem. Phys.* 135:065105.
12. Cohen, S. I. A., M. Vendruscolo, …, T. P. J. Knowles. 2011. Nucleated polymerization with secondary pathways. II. Determination of

self-consistent solutions to growth processes described by non-linear master equations. *J. Chem. Phys.* 135:065106.

13. Cohen, S. I. A., M. Vendruscolo, …, T. P. J. Knowles. 2011. Nucleated polymerization with secondary pathways. III. Equilibrium behavior and oligomer populations. *J. Chem. Phys.* 135:065107.
14. Morris, R. J., K. Eden, …, C. E. MacPhee. 2013. Mechanistic and environmental control of the prevalence and lifetime of amyloid oligomers. *Nat. Commun.* 4:1891.
15. Nielsen, L., R. Khurana, …, A. L. Fink. 2001. Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. *Biochemistry.* 40:6036–6046.
16. Padrick, S. B., and A. D. Miranker. 2002. Islet amyloid: phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis. *Biochemistry.* 41:4694–4703.
17. Gillam, J. E., and C. E. MacPhee. 2013. Modeling amyloid fibril formation kinetics: mechanisms of nucleation and growth. *J. Phys. Condens. Matter.* 25:373101.
18. Michaels, T. C. T., and T. P. J. Knowles. 2014. Role of filament annealing in the kinetics and thermodynamics of nucleated polymerization. *J. Chem. Phys.* 140:214904.
19. Hill, T. L. 1983. Length dependence of rate constants for end-to-end association and dissociation of equilibrium linear aggregates. *Biophys. J.* 44:285–288.
20. Knowles, T. P. J., D. A. White, …, D. A. Weitz. 2011. Observation of spatial propagation of amyloid assembly from single nuclei. *Proc. Natl. Acad. Sci. USA.* 108:14746–14751.
21. Oosawa, F., and S. Asakura. 1975. Thermodynamics of the Polymerization of Protein. Academic Press, New York.
22. Oosawa, F., and M. Kasai. 1962. A theory of linear and helical aggregations of macromolecules. *J. Mol. Biol.* 4:10–21.
23. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.
24. Morris, A. M., M. A. Watzky, …, R. G. Finke. 2008. Fitting neurological protein aggregation kinetic data via a 2-step, minimal/"Ockham's razor" model: the Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth. *Biochemistry.* 47:2413–2427.
25. Watzky, M. A., A. M. Morris, …, R. G. Finke. 2008. Fitting yeast and mammalian prion aggregation kinetic data with the Finke-Watzky two-step model of nucleation and autocatalytic growth. *Biochemistry.* 47: 10790–10800.
26. Cohen, S. I. A., M. Vendruscolo, …, T. P. J. Knowles. 2013. The kinetics and mechanisms of amyloid formation. *In* Amyloid Fibrils and Prefibrillar Aggregates. D. E. Otzen, editor. Wiley-VCH, Weinheim, Germany, pp. 183–209.
27. Wright, C. F., S. A. Teichmann, …, C. M. Dobson. 2005. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature.* 438:878–881.
28. Hellstrand, E., B. Boland, …, S. Linse. 2010. Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem. Neurosci.* 1:13–18.
29. Ramachandran, G., and J. B. Udgaonkar. 2012. Evidence for the existence of a secondary pathway for fibril growth during the aggregation of tau. *J. Mol. Biol.* 421:296–314.
30. Szavits-Nossan, J., K. Eden, …, R. J. Allen. 2014. Inherent variability in the kinetics of autocatalytic protein self-assembly. *Phys. Rev. Lett.* 113:098101.
31. Cohen, S., L. Rajah, …, T. Knowles. 2014. Spatial propagation of protein polymerization. *Phys. Rev. Lett.* 112:098101.
32. Serio, T. R., A. G. Cashikar, …, S. L. Lindquist. 2000. Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science.* 289:1317–1321.
33. Kelly, J. W. 2000. Mechanisms of amyloidogenesis. *Nat. Struct. Biol.* 7:824–826.
34. Collins, S. R., A. Douglass, …, J. S. Weissman. 2004. Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLoS Biol.* 2:e321.
35. Zhu, L., X.-J. Zhang, …, S. Perrett. 2003. Relationship between stability of folding intermediates and amyloid formation for the yeast prion Ure2p: a quantitative analysis of the effects of pH and buffer system. *J. Mol. Biol.* 328:235–254.
36. Foderà, V., F. Librizzi, …, M. Leone. 2008. Secondary nucleation and accessible surface in insulin amyloid fibril formation. *J. Phys. Chem. B.* 112:3853–3858.
37. Bucciantini, M., G. Calloni, …, M. Stefani. 2004. Prefibrillar amyloid protein aggregates share common features of cytotoxicity. *J. Biol. Chem.* 279:31374–31382.
38. Xue, W.-F., A. L. Hellewell, …, S. E. Radford. 2009. Fibril fragmentation enhances amyloid cytotoxicity. *J. Biol. Chem.* 284:34272–34282.