Electronic Supplementary Material (ESI) for Chemical Science.

Supporting information: Kinetic models reveal the interplay of protein production and aggregation

Jiapeng Wei,[†] Georg Meisl,[†] Alexander Dear,[†,‡] Matthijs Oosterhuis,[¶,§] Ronald Melki,[‖] Cecilia Emanuelsson,[§] Sara Linse,[⊥] and Tuomas P. J. Knowles[*,†,#]

† *Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK*

‡ *Department of Biochemistry and Structural Biology, Lund University, SE22100 Lund, Sweden*

¶*Current address: Patheon Biologics, Zuiderweg 72/2, 9727 DL Groningen*

§*Department of Biochemistry and Structural Biology, Center for Molecular Protein Science,Lund University, Sweden*

‖*Dynamique du Cytosquelette, Laboratoire d'Enzymologie et Biochimie Structurales, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette Cedex, France*

⊥*Department of Biochemistry and Structural Biology, Lund University, Lund, Sweden*

#*Cavendish Laboratory, University of Cambridge, J J Thomson Avenue, CB3 0HE, UK*

E-mail: tpjk2@cm.ac.uk(Tuomas.P.J.Knowles)

1 Methods

1.1 Chemicals and materials

All chemicals were of analytical grade unless specified. Buffers were filtered (200 nm filter) and degassed before use. Thioflavin T (ThT) was from CalBiochem and a 2 mM stock solution was prepared from ThT powder, dissolved in Millipore water, passed through a 200 nm filter, and the concentration determined from the absorbance of diluted stock samples. The PEGylated black polystyrene 96-well half-area plates with clear bottom were from Corning (3881).

1.2 Protein purification

The His6-tagged fusion protein of maltose-binding protein (MBP) and Htt were expressed and purified as described.[1]

1.3 Measurement of enzymatic cleavage

The enzymatic cleavage reaction was studied for MBP-HttQ30 at six concentrations ranging from 6.4 to 20 μM with 0.54 μM TEV protease in buffer (20 mM Tris-HCl, pH 7.5, 125 mM KCl). Samples were incubated at 37 °C with aliquots withdrawn at different time points and separated by SDS PAGE (NuPAGE 4-12% Bis-Tris Gel, provider ThermoFisher, Stockholm, Sweden) using an image scanner for assessment of the amounts of cleaved and un-cleaved MBP-HttQ30.

1.4 Measurement of kinetics

Aggregation kinetics were monitored at 37 °C via the ThT fluorescence intensity measured continuously using a BMG Fluostar plate reader with an excitation filter at 448 nm and emission filter at 480 nm. Purified MBP-HttQ45 at 4-15 μM total concentration in buffer (20 mM Tris-HCl, pH 7.5, 125 mM KCl) with 10 μM ThT, without or with preformed seeds fibrils, was placed in the wells of PEGylated polystyrene plates. TEV protease (AcTEV Protease, provider ThermoFisher, Stockholm, Sweden) was added at time zero to a final concentration of maximum 0.54 μM, in a dilution series with activity varying between 7 U and 0.25U, after which the reading of ThT fluorescence intensity was immediately started.

1.5 Fitting of kinetic data

All fits were performed on the modified local version of fitting platform AmyloFit.[2] The differential equations describing the kinetics and their approximate solutions which were used in the fitting are all given in the main text and ESI [†].

2 The fundamentals of aggregation without a source-term

When primary nucleation, elongation and secondary nucleation are involved in the aggregation of a protein, the time-evolution of the aggregate distribution $f(t, j)$, of aggregates

of size j at time t, is given by:[3]

$$\begin{aligned}\frac{\partial f(t,j)}{\partial t} &= k_n m(t)^{n_c}\delta_{j,n_c} && \text{(Primary nucleation)}\\ &+ 2k_+ m(t)^{n_+}[f(t,j-1)-f(t,j)] && \text{(Elongation)}\\ &+ k_2 m(t)^{n_2}\sum_{i=n_c}^{\infty} i f(t,i)\delta_{j,n_2} && \text{(Secondary nucleation).}\end{aligned} \tag{S1}$$

where $m(t)$ is the free monomer concentration. k_n, k_2, k_+ are the rate constants of the primary nucleation, secondary nucleation and elongation. n_c, n_2 are reaction orders of primary nucleation and secondary nucleation, which in the simplest interpretation correspond to nucleus size. n_+ is the reaction order of elongation, which similarly can be interpreted as the number of monomers that are added onto the growing fibril in a single elongation step. In most cases, elongation happens by monomer addition so $n_+ = 1$. $\delta_{i,j}$ is the Kronecker delta which equals 1 if $i = j$ and 0 otherwise.

As a set of infinitely many coupled and non-linear differential equations, the master equations Eqs. (S1) are usually be very difficult to deal with. However, the average quantities such as total fibril number and total fibril mass are both easier to measure and more mathematically tractable. The principal moments of the aggregation length distribution are defined as:

$$Q_N(t) = \sum_{j=n_c}^{\infty} j^N f(t,j). \tag{S2}$$

We usually use number of aggregates, the zero-th moment $P(t) = Q_0(t) = \sum_{j=n_c}^{\infty} f(t,j)$, and the total mass of aggregates, the first moment $M(t) = Q_1(t) = \sum_{j=n_c}^{\infty} j f(t,j)$, to describe an aggregation reaction.[4–6] Combining Eqs. (S1) and (S2), we obtain the

moment equations:[7–9]

$$\frac{dP(t)}{dt} = k_n m(t)^{n_c} + k_2 m(t)^{n_2} M(t) \tag{S3a}$$

$$\frac{dM(t)}{dt} = 2k_+ m(t)^{n_+} P(t) \tag{S3b}$$

$$m(t) + M(t) = m_{tot}. \tag{S3c}$$

where m_{tot} is total protein concentration of the system and Eqs. (S3c) describes the conservation law between free monomers and the fibril mass, under the assumption that no more new monomers are added into the system during the aggregation.[10,11] The moment equations Eqs. (S3) can be solved for example using a fixed point iteration method:[7] Firstly, approximate $m(t) = m_{tot}$ to enable solving the resulting differential equations and obtain an approximate early time solution. Secondly, iterate the early time solution into the moment equations to obtain the approximate full timescale solution.

3 Compare source-term aggregation with aggregation without secondary nucleation

For the source-term $s(t) = m_{tot}(1 - e^{-k_{source}t})$, Eqs. (1) can be numerically solved and the normalized fibril mass under different conditions are plotted in Fig. S1: A and B compare the aggregation with secondary nucleation ($k_2 \neq 0$) but without the source-term ($s(t)/m_{tot} = 1$) to the aggregation with the source-term $s(t)/m_{tot} = 1 - e^{-k_{source}t}$ but without secondary nucleation ($k_2 = 0$). Although the early time solution of the aggregation with the source-term but without secondary nucleation is a polynomial function: $M_0(t) = \frac{k_+ k_n m_{tot}^3}{6k_{source}^2}((6k_{source}^2 t^2 - 18k_{source}t + 17) + 6e^{-k_{source}t}(2k_{source}t - 5) + 15e^{-2k_{source}t} - 2e^{-3k_{source}t})$, while the early time solution of the aggregation without a source-term but with the sec-

ondary nucleation is an exponential function: $M_0(t) = \frac{k_n}{k_2} m_{tot}^{n_c - n_2} Sinh^2(\frac{\sqrt{2k_+ k_2 m_{tot}^{n_2+1}}}{2} t)$, their shape can be very similar at the condition of a certain set of parameters, as shown in Fig. S1A. However, these two curves can be distinguished by adding seeds at $t = 0$. For aggregation with a source-term but without the secondary nucleation, after adding seeds, only the elongation speed is increased at the early time and the aggregation speed is still mainly constrained by the source-term. For aggregation without a source-term but with the secondary nucleation, adding seeds at $t = 0$ bypasses the primary nucleation process, largely increases the speed of the surface catalyzed secondary nucleation at the early time and the aggregated fibril mass rises earlier, as shown in Fig. S1B. Adding 10% seeds fibrils with the average length of 100 can distinguish these two types of aggregations.

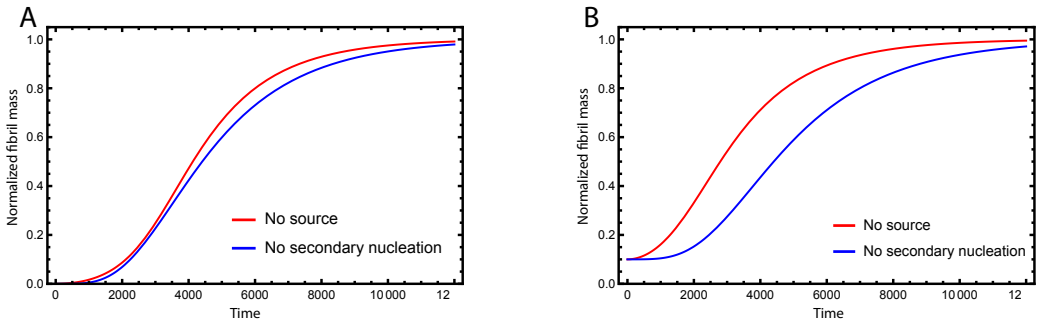

Figure S1: Numerical solution of normalized fibril mass concentration $M(t)/m_{tot}$. A and B compare the aggregation with secondary nucleation but without the source-term (red curves) and the aggregation with the source-term $s(t) = 1 - e^{-k_{source}t}$ but without secondary nucleation (blue curves). Both aggregations in A are unseeded. Both aggregations in B have seeded fibrils initially: $M(0) = 0.1 \cdot m_{tot}, P(0) = 0.01 \cdot M(0)$. $n_c = n_2 = 2$ for all panels.

4 Using time dependent effective parameters to analytically solve aggregation with a source-term

Writing the inverse of $S(T)$ as $T = T(S)$, we can rewrite the early-time kinetic Eqs. (4) as an ODE with independent variable S. In Sturm-Liouville form[12] this is:

$$[\frac{G(S)}{S^{n_+}} \cdot \nu_0'(S)]' = 2\epsilon \cdot \frac{S^{n_c}}{G(S)} + \nu_0(S) \cdot \frac{S^{n_2}}{G(S)}, \tag{S4}$$

where $G(S) = \frac{dS}{dT}$. This can be analytically solved when $n_c = n_2 = n_+ = n_0$ (n_0 is defined as a 'homogeneous' reaction order for all reaction micro-steps), yielding:

$$\hat{\nu_0}(S) = \frac{1}{2}e^{-H(S)}[(e^{H(S)} - 2\epsilon)^2 - 2c_1], H(S) = \int_S^1 \frac{x^{n_0}}{G(x)}dx + c_2. \tag{S5}$$

where c_1, c_2 are decided by boundary conditions. For the source-term $S(T) = 1 - e^{-KT}$, Eqs. (S5) becomes:

$$\hat{\nu_0}(S) = 4\epsilon \cdot \sinh^2\left[\frac{\sum_{n=1}^{n_0}(\frac{S^n}{n}) + \ln(1-S)}{2K}\right]. \tag{S6}$$

In general, $n_+ = 1$; typically, however, the reaction orders $n_c, n_2 \neq 1$. Nonetheless, we can obtain an approximate analytical solution of Eqs. (4) for arbitrary n_c, n_2 based on the special solution Eq. (S6) by introducing time dependent effective parameters. The time dependent effective parameters are used to correct the approximate analytical solution from the 'homogeneous' reaction order $n_c = n_2 = n_+ = n_0$ to the more general 'inhomogeneous' reaction orders. To see how this works, we rewrite Eqs. (4) as:

$$\Pi_0'(T) = 2\epsilon S^{n_c-n_0} S(T)^{n_0} + S^{n_2-n_0}\nu_0(T)S(T)^{n_0} \tag{S7a}$$

$$\nu_0'(T) = S^{n_+-n_0} S(T)^{n_0}\Pi_0(T). \tag{S7b}$$

When S is slowly varying compared to ν_0 and Π_0, i.e. when monomer production is either slow or rapid relative to aggregation, we have timescale separation and can to a good approximation treat $S^{n_x-n_0}$ as constants. Thus, defining $\bar{T} = \kappa_{\text{eff}} t = \bar{\kappa} T$, and $\bar{\Pi}_0 = S^{n_+-n_0}\Pi_0/\bar{\kappa}$, Eq. (S7b) becomes:

$$\nu_0'(\bar{T}) = S(\bar{T})^{n_0}\bar{\Pi}_0(\bar{T}). \tag{S8}$$

Moreover, Eq. (4a) becomes:

$$\frac{\bar{\kappa}^2}{S^{n_+-n_0}}\bar{\Pi}_0'(\bar{T}) = 2\epsilon S^{n_c-n_0} S(\bar{T})^{n_0} + S^{n_2-n_0}\nu_0(\bar{T})S(\bar{T})^{n_0} \tag{S9}$$

$$\bar{\Pi}_0'(\bar{T}) = 2\epsilon^{\text{eff}} S(\bar{T})^{n_0} + \nu_0(\bar{T})S(\bar{T})^{n_0} \tag{S10}$$

$$\bar{\kappa}^2 = S^{n_+ + n_2 - 2n_0}, \quad \epsilon^{\text{eff}} = \epsilon \cdot S^{n_c-n_2} \tag{S11}$$

$$S(\bar{T}) = 1 - e^{-K^{\text{eff}}\bar{T}}, \quad K^{\text{eff}} = \frac{K}{\bar{\kappa}} = K \cdot S^{\frac{2n_0-n_2-n_+}{2}}. \tag{S12}$$

The solution to this is Eq. (S6) with K replaced by K^{eff}, and ϵ replaced by ϵ^{eff}. Setting $n_+ = 1$, this yields:

$$\nu_0(S) = 4\epsilon \cdot S^{n_c-n_2} \cdot Sinh^2\left[\frac{\sum_{n=1}^{n_0}(\frac{S^n}{n}) + \ln(1-S)}{2K \cdot S^{\frac{2n_0-n_2-1}{2}}}\right]. \tag{S13}$$

The 'homogeneous' reaction order n_0 can be chosen as $n_0 = n_2$ if n_2 is a integer and $n_0 = \lceil n_2 \rceil$ otherwise. Then Eqs. (S13) is same as Eqs. (5).

5 Approximate early time solution in different extreme limits

We can simplify Eqs. (5) under different extreme limits:

For slow source, we can apply Taylor expansion on $\ln(1-S) \approx -S - \frac{S^2}{2} - \frac{S^3}{3} - ... - \frac{S^{n_2}}{n_2} - ....$ In this condition, the early time solution is:

$$\nu_0(t) \approx 4\epsilon \cdot s(t)^{n_c - n_2} \cdot Sinh^2[\frac{\kappa \cdot s(t)^{\frac{n_2+3}{2}}}{2(n_2+1)k_{source}}]. \tag{S14}$$

For secondary nucleation dominated, we can approximate Eqs. (S14) to:

$$\nu_0(t) \approx \epsilon \cdot s(t)^{n_c - n_2} \cdot e^{\frac{\kappa \cdot s(t)^{\frac{n_2+3}{2}}}{(n_2+1)\cdot k_{source}}}. \tag{S15}$$

For primary nucleation dominated, we can approximate Eqs. (S14) to:

$$\nu_0(t) \approx \frac{\lambda^2 \cdot s(t)^{n_c+3}}{(n_2+1)^2 \cdot k_{source}^2}. \tag{S16}$$

For fast source, we can approximate that $S \approx 1$. In this condition, the early time solution is:

$$\nu_0(t) \approx 4\epsilon \cdot Sinh^2(\frac{\kappa t}{2}). \tag{S17}$$

For secondary nucleation dominated, we can approximate Eqs. (S17) as:

$$\nu_0(t) \approx \epsilon \cdot e^{\kappa t}. \tag{S18}$$

For primary nucleation dominated, we can approximate Eqs. (S17) to:

$$\nu_0(t) \approx \frac{\lambda^2 t^2}{2}. \tag{S19}$$

There approximated results are listed in Table 1.

6 Using Lie symmetry to get the full timescale solution

To develop a global approximate solution we employ the method of asymptotic Lie symmetries. For explanation of this method, see ref.[13] Its applicability here requires that the structure in phase space of the dynamics be the same as in 'ordinary" protein aggregation kinetics without a source term, i.e. that μ and Π monotonically decrease and increase respectively from their initial values to an attractive fixed point at $\mu = 0,\ \Pi = \Pi_\infty$. Clearly this is not the case for Eqs. (2), since μ increases from zero initially. So, we must first transform the variables into a form that ensures the correct phase space structure. The logical way to do this is to divide through by S, i.e. $\bar{\mu} = \mu/S$, and $\bar{\nu} = \nu/S$. Eqs. (2) become:

$$\Pi'(T) = 2\epsilon\bar{\mu}(T)^{n_c} S^{n_c} + (1-\bar{\mu})\mu(T)^{n_2} S^{n_2+1} \tag{S20a}$$

$$\mu'(T) = S' - \mu(T)^{n_+}\Pi(T) \tag{S20b}$$

$$\bar{\mu}'(T) = -\bar{\mu}(T)^{n_+} S^{n_+-1}\Pi(T) - \frac{S'}{S}(\bar{\mu}-1) \tag{S20c}$$

$$1 = \bar{\mu}(T) + \bar{\nu}(T) \tag{S20d}$$

$$\Pi(0) = 0, \quad \bar{\mu}(0) = 1. \tag{S20e}$$

6.1 Perturbative solution

We now seek a perturbative solution by expanding $\bar{\mu}, \Pi$ in ϵ as $\bar{\mu} = 1 + \epsilon\bar{\mu}^{(1)}$ and $\Pi = \Pi^{(0)} + \epsilon\Pi^{(1)}$. At zeroth order, Eqs. (S20) then reduce to $(\Pi^{(0)})' = 0$. Given initial condition $\Pi(0) = 0$, this clearly is solved by $\Pi^{(0)} = 0$, the same as with no source term.[13]

At first order in ϵ, Eqs. (S20) are:

$$(\Pi^{(1)}(T))' = 2S^{n_c} - \bar{\mu}^{(1)}S^{n_2+1} \tag{S21a}$$

$$(\bar{\mu}^{(1)}(T))' = -S^{n_+-1}\Pi^{(1)}(T) - \frac{S'}{S}\bar{\mu}^{(1)} \tag{S21b}$$

$$\therefore (S\bar{\mu}^{(1)}(T))' = -S^{n_+}\Pi^{(1)}(T). \tag{S21c}$$

Rewriting for $\mu^{(1)} = S\bar{\mu}^{(1)}$, this yields:

$$\frac{d\Pi^{(1)}}{dT} = 2S^{n_c} - \mu^{(1)}S^{n_2}, \quad \frac{d\mu^{(1)}}{dT} = -S^{n_+}\Pi^{(1)}. \tag{S22}$$

Adding this to the zeroth-order perturbation equations and making the substitution $\nu_0 = -\mu^{(1)}$ yields the "early-time" equations Eqs. (4), demonstrating their equivalence. These have been solved for ν_0 in Appendix 4, yielding Eqs. (5). So, for $\bar{\mu}^{(1)}$, we have:

$$\bar{\mu}^{(1)} = -4S^{n_c-n_2-1} \cdot \sinh^2\left[\frac{\sum_{n=1}^{n_2}\frac{S^n}{n} + \ln(1-S)}{2K \cdot S^{(n_2-1)/2}}\right]. \tag{S23}$$

6.2 Lie symmetry solution

In the limit that $S \to 1$ (or $K \to \infty$), the kinetic equations recover the traditional protein aggregation equations. A $\mu \to 1$ asymptotic symmetry connecting S with ϵ and T should suffice to transform the solution for μ in protein aggregation with no source

into a solution for $\bar{\mu}$ valid for aggregation with a source term. This follows since $\bar{\mu} \to \mu$ in the limit $S \to 1$, and since we have shown that the late-time kinetics are identical and unaffected by S to leading order. The source-less solution is:[13]

$$\mu = \left(1 + 4\epsilon \sinh^2(T/2)/c\right)^{-c}, \tag{S24}$$

where $c = 3/(2n_2 + 1)$.

As discussed in ref.,[13] it is not necessary to explicitly compute the symmetry; we can instead simply replace ϵ and T in the ordinary protein aggregation solution with functions of S that ensure that its series expansion in ϵ matches the perturbative solution Eq. (S23). This follows because a condition of the asymptotic symmetry transformation of these parameters is that the series expansion matches.

In the present case, this requires $\epsilon \to \epsilon S^{n_c - n_2 - 1}$, and $T \to (\sum_{n=1}^{n_2} \frac{S^n}{n} + \ln(1-S))/(K \cdot S^{(n_2-1)/2})$. This finally yields the general solution:

$$\bar{\nu} = 1 - (1 + \epsilon\bar{\nu}_1/c)^{-c}. \tag{S25}$$

This general solution is Eqs. (6).

Substituting the early time solution Eqs. (5) into the unified equation Eqs. (6), we obtain the approximate full timescale solution for aggregation with a source-term. This approximate analytical solution converges on the exact numerical kinetics almost precisely, as shown in Fig. S2.

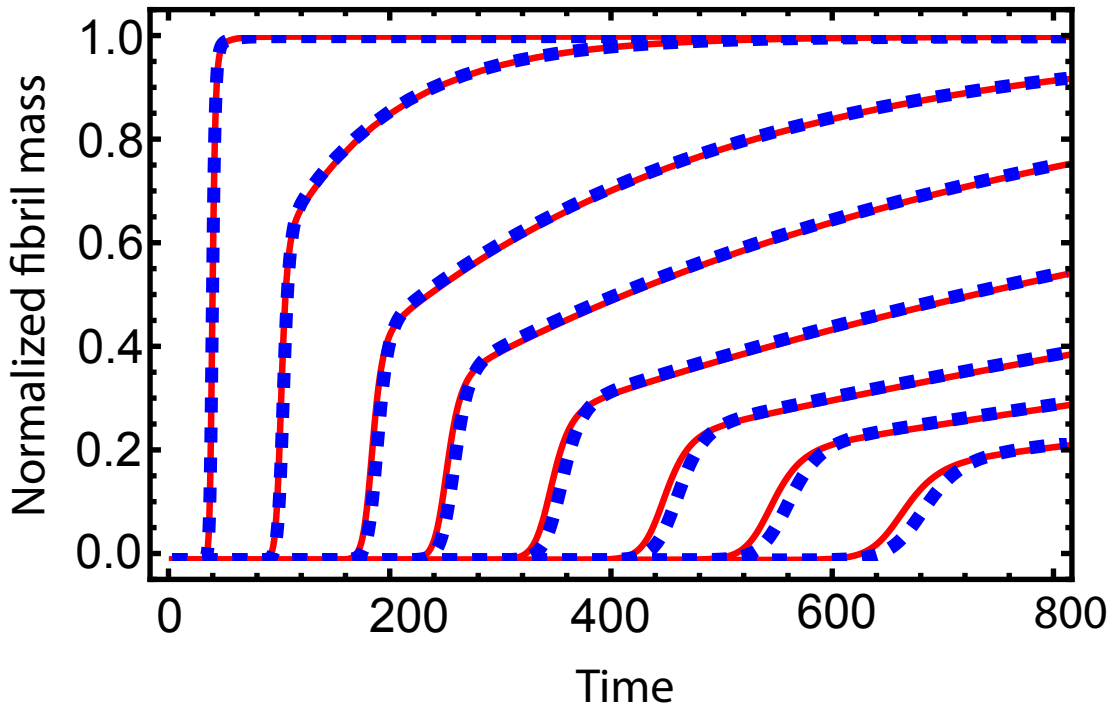

Figure S2: Comparing the analytical solution with the numerical integrated rate law. Blue dotted curves are the approximate analytical solutions Eqs. (6). Red curves are the numerically integrated rate law Eqs. (2)-(3). From right to left, the parameters are: $K = 10^{-3.5}, 10^{-3.35}, 10^{-3.2}, 10^{-3}, 10^{-2.75}, 10^{-2.5}, 10^{-2}, 10^{-1}$. $n_c = n_2 = 2$ and $\epsilon = 10^{-10}$ for all curves.

7 Gel images of enzyme cleavage

Fig. S3 A-B show the degree of enzymatic cleavage of MBP-HttQ45.

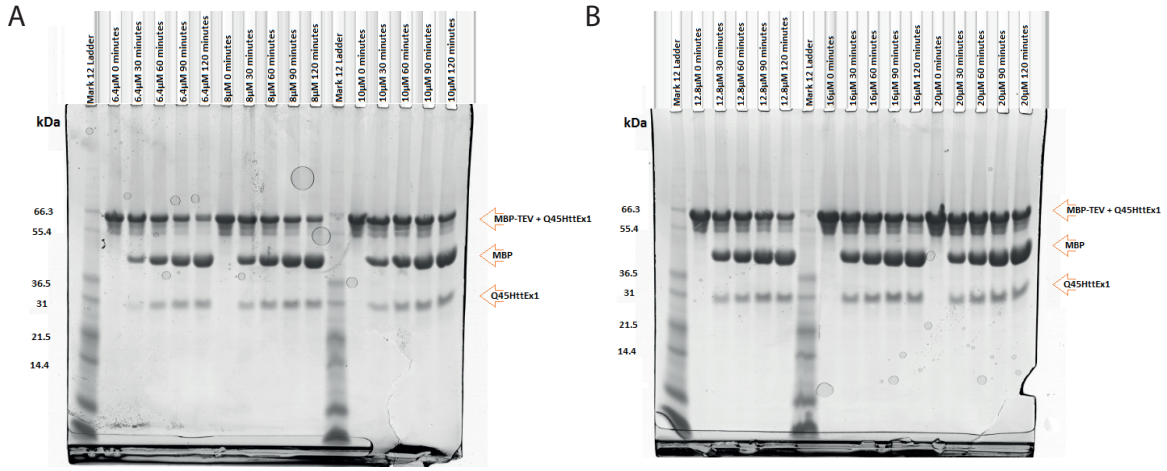

Figure S3: Gel images of enzyme cleavage. Gel electrophoresis images taken at $0-120$ minutes. The initial concentrations of MBP-HttQ45 are: 6.4μM, 8μM, $10\mu Min$ panel A and 12.8μM, 16μM, 20μM in panel B.

8 Bayesian analysis in enzyme cleavage data fitting

We global fit the enzyme cleavage data (with different m_{tot} but the same $e_0 = 0.54\mu$M) to the source-term: $s(t) = m_{tot} - \frac{k_b+k_c}{k_a} \cdot W_0[\frac{k_a}{k_b+k_c} \cdot m_{tot} \cdot e^{\frac{k_a}{k_b+K_c}(m_{tot}-k_c e_0 t)}]$.

We use Bayesian analysis to find the optimised k_a, k_b, k_c. If we note the measured data points (t_i, s_i) as d and the fitting parameters k_a, k_b, k_c as k, the probability of fitting parameters k in the condition of data points d, the likelihood $P(k|d)$ can be expressed in Bayesian theorem as follows:

$$P(k|d) = \frac{P(d|k)P(k)}{P(d)}. \tag{S26}$$

where $P(d|k)$ is the posterior probability, $P(k)$ is the prior probability and $P(d)$ is the marginal likelihood (usually is a constant). In order to find the optimized fitting parameters, we need to find k_a, k_b, k_c which leads to the maximum of the likelihood $P(k|d)$. If we assume that we have a flat prior probability, which implies $P(k)$ is a constant, $P(k|d) \propto P(d|k)$. We can find k_a, k_b, k_c that leads to the maximum of the posterior probability $P(d|k)$. We assume the posterior probability is a normal distribution and multiplies $P_i(d|k)$ together for all data points d_i:

$$\prod_i^n P(d|k) = (\frac{1}{\sigma\sqrt{2\pi}})^n \cdot e^{-\frac{1}{2\sigma^2}\sum_i^n [s(t_i,k)-s_i]^2}. \tag{S27}$$

where σ is the averaged standard deviation of all s_i calculated from the triplicate measurements.

The 1D and 2D projections of multiplication posterior probability are shown in Fig. S4. In 2D plot, brighter color represents larger posterior probability and more optimised fitting data k. In 1D plot, larger value represents more optimised fitting data k. For $Q45$ enzyme cleavage measurements, we fit that $k_a \approx 0.0355$, and $k_a \ll k_c$. Substituting them into: $s(t) = m_{tot} - \frac{k_b+k_c}{k_a} \cdot W_0[\frac{k_a}{k_b+k_c} \cdot m_{tot} \cdot e^{\frac{k_a}{k_b+K_c}(m_{tot}-k_c e_0 t)}]$. For $k_a \ll k_c$, we can approximate the fitting function: $s(t) \approx m_{tot}(1-e^{-k_{source}t})$, where $k_{source} = \frac{k_a k_c e_0}{k_b+k_c}$. Using this approximation, we can calculate that $k_{source}^{Q45} = 0.02 min^{-1}\mu\mathrm{M}^{-1}$ (from the optimised k_a, k_b, k_c fitted by Bayesian method).

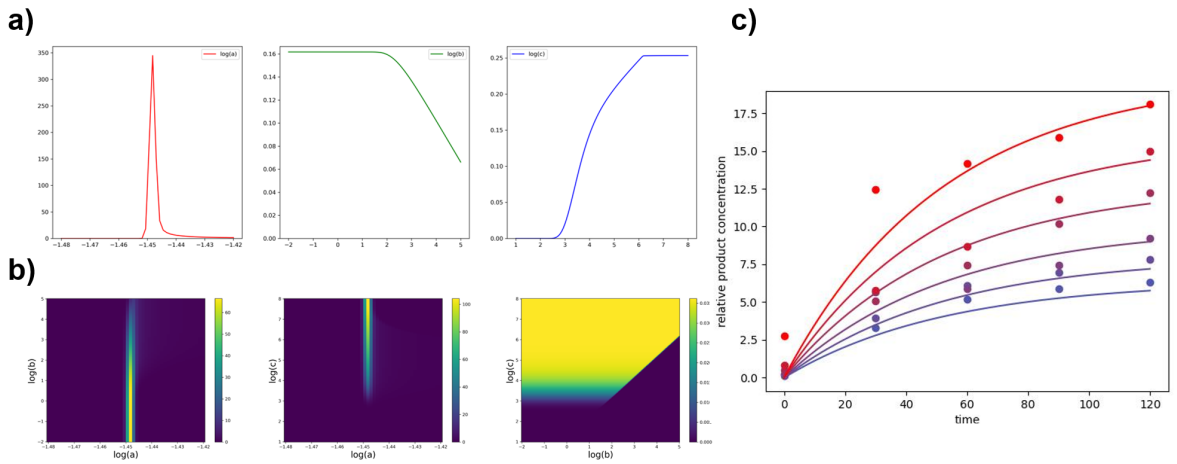

Figure S4: 1D and 2D projections multiplication of posterior probability of $Q45$ data set are shown in a) and b). c) is the fitting curves plotted using the optimised k_a, k_b, k_c from a) and b). $m_{tot} = 6.4\mu\text{M}, 8\mu\text{M}, 10\mu\text{M}, 12.8\mu\text{M}, 16\mu\text{M}, 20\mu\text{M}$, $e_0 = 0.54\mu\text{M} \ll m_{tot}$.

9 Add additional TEV at 24h after HttQ45 aggregation

The aggregation kinetic of seeded HttQ45 aggregation with very low level of enzyme in the beginning but add a lot of enzyme after 24h is shown in Fig. S5: We can see the plateaued ThT signal rises again after adding additional plenty of enzyme, which implies that the firstly added enzyme has been somehow consumed and there is still non-aggregating precursor remains in the system.

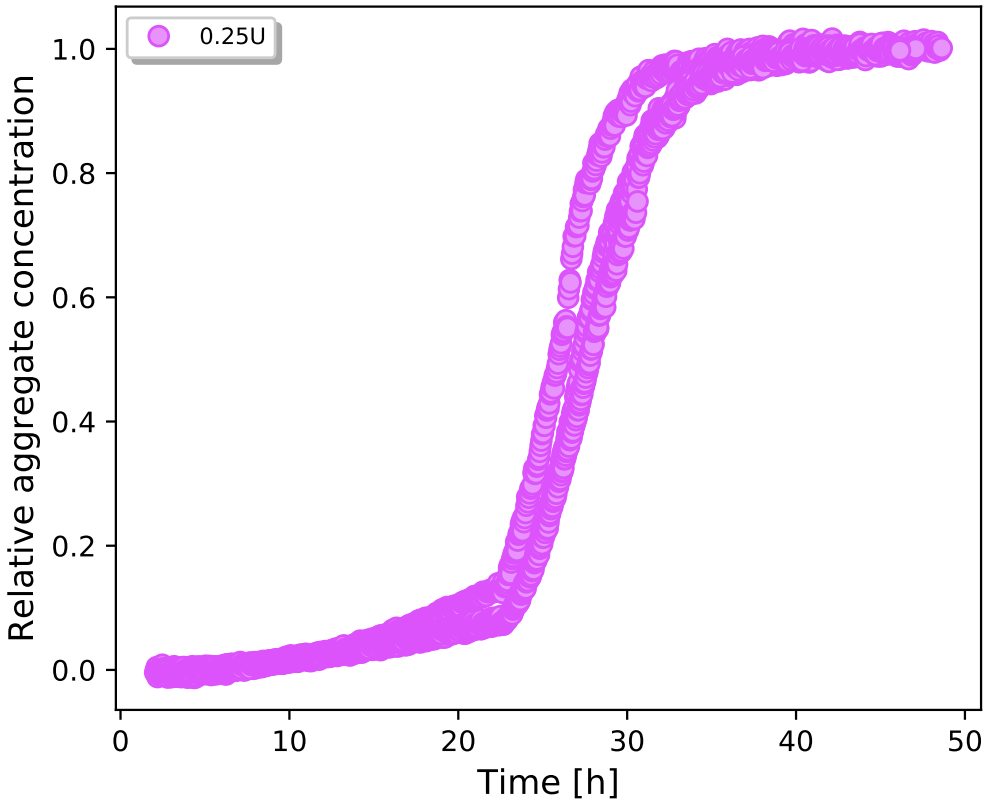

Figure S5: Beginning with 8μM monomer, 30% seeding, 0.003/muM enzyme. Further add 0.1μM of enzyme after 24 h.

10 Same concentration of enzyme with different concentration of monomer

For the same concentration of enzyme and different concentrations of monomers, larger m_{tot} causes smaller percentage of non-aggregating precursors can finally aggregate, as shown in Fig. S6:

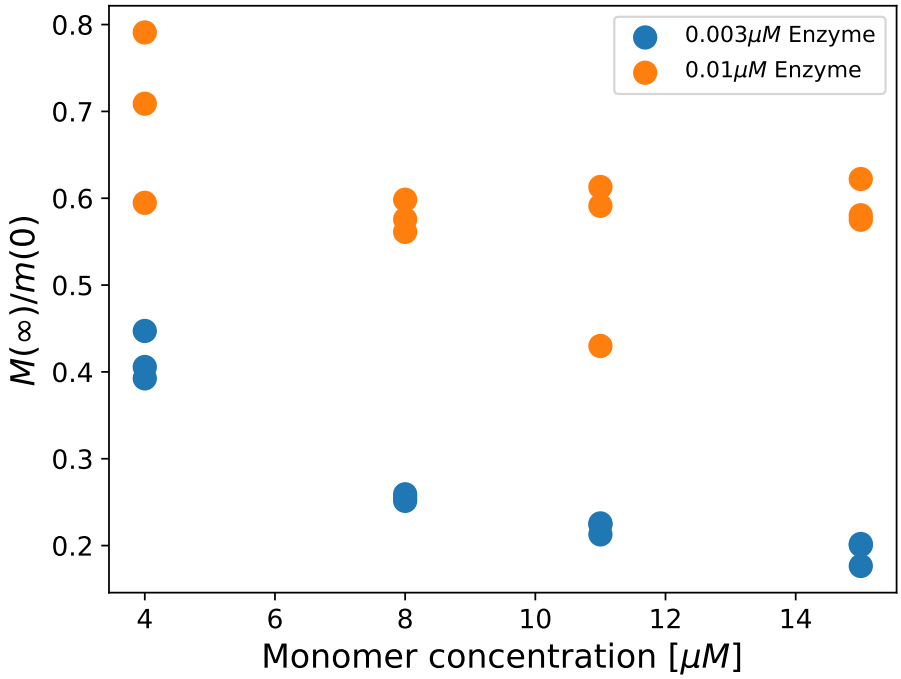

Figure S6: 4μM, 8μM, 11μM and 15μM monomer with 0.01μM and 0.003μM enzyme. Use the ThT signal to justify $M(\infty)/m(0)$. The value of $m(0)$ can be calibrated by the ThT signal of same monomer concentration but with sufficient enzyme.

11 A simple source-term model for in-vivo protein expression and clearance

As detailed in the main text, we assume aggregation-prone monomers are generated at a constant production rate, with rate constant k_p and removed at a rate proportional to the concentration and rate constant k_r. This produces the following moment equations

as:

$$\frac{dP(t)}{dt} = k_n m(t)^{n_c} + k_2 m(t)^{n_2} M(t) \tag{S28a}$$

$$\frac{dM(t)}{dt} = 2k_+ m(t)^{n_+} P(t) \tag{S28b}$$

$$\frac{dm(t)}{dt} = k_p - k_r m(t) - 2k_+ m(t)^{n_+} P(t) \tag{S28c}$$

$$M(0) = m(0) = P(0) = 0. \tag{S28d}$$

The numerical solution of Eqs. S28 is shown in Fig. 6. $M(t)$ is the blue curve and $m(t)$ is the red curve.

Under the early time approximation, the aggregation consumes very few monomers. We can approximate Eqs. S28c as:

$$\frac{dm(t)}{dt} \approx k_p - k_r m(t), \quad m(0) = 0. \tag{S29}$$

We solve $m(t)$ and get the source-term:

$$s(t) = m(t) + M(t) = \frac{k_p}{k_r}(1 - e^{-k_r \cdot t}). \tag{S30}$$

where $M(t) = 0$ for the early time.

In order to get the early time analytical solution, we substitute the source-term into Eqs. 5 and get:

$$M_0(t) = 2\frac{k_n}{k_2}(\frac{k_p}{k_r})^{n_c - n_2}(1 - e^{-k_r t})^{n_c - n_2} \sinh^2[\frac{-k_r t + (1 - e^{-k_r t}) + \frac{(1-e^{-k_r t})^2}{2}}{\frac{2k_r}{\sqrt{2k_+ k_2 (\frac{k_p}{k_r})^{n_2+1}}}(1 - e^{-k_r t})^{\frac{n_2-1}{2}}}]. \tag{S31}$$

which is plotted as the green curve in Fig. 6A. Compared with the numerical blue curve,

we find that the approximate early time solution Eqs. S31 describes the lag time well.

We further substitute the early time solution Eqs. S31 into Eqs. 6 to get the full time solution. At late time, the aggregation-prone monomer concentration approaches the peptide solubility due to a large amount of fibrils that have been generated. The increase in aggregate mass at late time thus only depends on the monomer production rate. The late time solution in Eqs. 6 is $M_\infty(t) = k_p t$. We plot the analytical full time solution as the black curve in Fig. 6A. We also plot the numerical aggregate mass and the analytical mass at a longer time scale, as shown in Fig. 6B. We find that the exponential growth of fibril mass only appears at the early time and at the late time the fibril mass grows linearly due to the limit of the constant monomer production rate.

12 Full time solutions with and without source term

We substitute the early time solution from Table. 1 column 4 into Eqs. 6 to get the full time solutions for all regimes.

13 Boundary line between the slow monomer production region and the fast monomer production region

If we take the logarithm of both sides of the lag time solution in Table 1 regions 1 and 3 and cancel the t_{lag}, we obtain the approximate boundary line between the slow monomer production region and the fast monomer production region:

$$\log_{10}\kappa \approx \log_{10} k_{source} + \log_{10}(\ln\frac{1}{\epsilon}) - \frac{2}{n_2+1}\log_{10}(n_0+1). \tag{S32}$$

Table S1: Full time solutions in five different regimes.

Regime	Dominant nucleation	Source	Full time solutions $M(t)$
1	Secondary nucleation	Slow source	$(1-(1+\frac{\epsilon\cdot s(t)^{n_c-n_2-1}\cdot e^{\frac{\kappa\cdot s(t)^{\frac{n_2+3}{2}}}{(n_2+1)\cdot k_{\text{source}}}}}{\frac{3}{2n_2+1}})^{-\frac{3}{2n_2+1}})\cdot s(t)$
2	Primary nucleation	Slow source	$(1-(1+\frac{(2n_2+1)\cdot\lambda^2\cdot s(t)^{n_c+2}}{3(n_2+1)^2\cdot k_{\text{source}}^2})^{-\frac{3}{2n_2+1}})\cdot s(t)$
3	Secondary nucleation	Fast source	$(1-(1+\frac{\epsilon\cdot e^{\kappa t}\cdot(2n_2+1)}{3})^{-\frac{3}{2n_2+1}})\cdot m_{tot}$
4	Primary nucleation	Fast source	$(1-(1+\frac{\lambda^2t^2(2n_2+1)}{6})^{-\frac{3}{2n_2+1}})\cdot m_{tot}$
5	N.A.	Very slow source	$s(t)$

The region boundary is roughly a straight line in the lag time contour plot. The slope of the boundary line is 1. The intercept of the boundary line is $\log_{10}(\ln\frac{1}{\epsilon})-\frac{2}{n_2+1}\log_{10}(n_0+1)$. The first term of the intercept $\log_{10}(\ln\frac{1}{\epsilon})$ depends on the y-axis $\log_{10}\kappa$. We can approximate it as a constant since ϵ is in a nested logarithm's function. For $10^{-14}<\epsilon<10^{-6}$, the first term of the intercept is $1.14<\log_{10}(\ln\frac{1}{\epsilon})<1.5$. For $n_0=1,2,3,4,5$ and $0<n_2<5$, the second term of the intercept is $0.26<\frac{2}{n_2+1}\log_{10}(n_0+1)<0.32$. We can sum these two terms and approximate the intercept of the boundary line equals to 1. The boundary line in the lag time contour plot (double-log) is shown in Fig. S7 as the white line. We find that the analytical prediction of the region boundary $\log_{10}\kappa\approx\log_{10}k_{source}+1$ fit well with the numerical fitting in Fig. S7.

By using the same method, we can further calculate all boundaries between these five regions, as shown in Table S2.

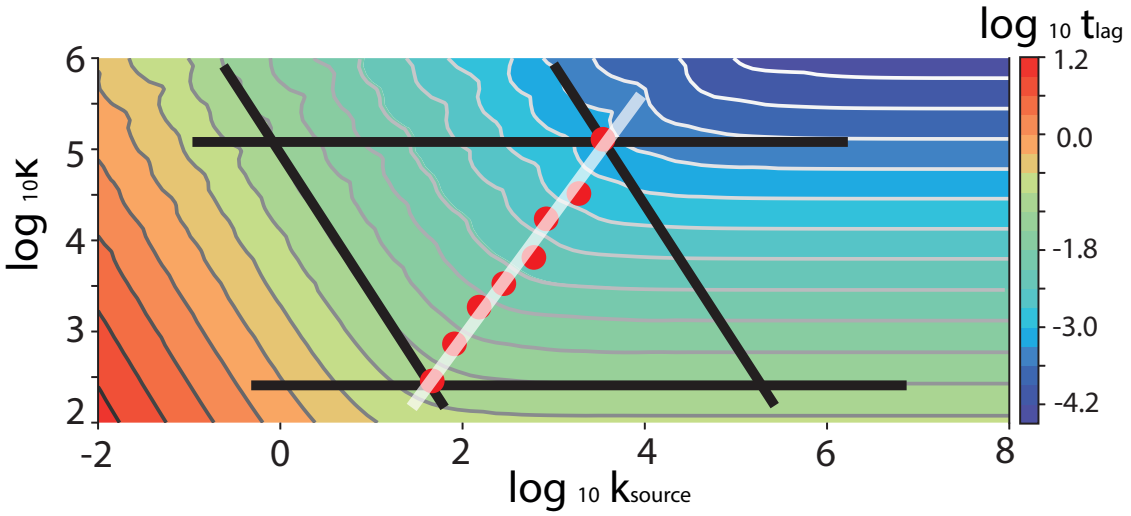

Figure S7: Contour plot of exact numerical lag time when $n_c = n_2 = 2, n_+ = 1$. Two sets of black parallel lines are the extension of the lag time contour lines of the slow monomer production region and the fast monomer production region. Red points are the cross points of the corresponding extension black lines. These red points form our numerical estimates of the boundary line. The analytic boundary line between these two regions shown as a 45° white line (fitted by the red cross points): $\log_{10}\kappa = \log_{10} k_{source}+1$. $\lambda = 0.1$. The value of λ leads to $10^{-14} < \epsilon < 10^{-6}$ when $10^2 < \kappa < 10^6$.

14 Reversible source term solutions

Having discussed the general features of a class of source terms that can be well approximated by first order irreversible kinetics, we now explore how more complex source terms alter the system behaviour. First, we consider the reversibility of the monomer production reaction, which may be important in a number of systems. For instance, when the source consists of non-aggregating precursors unfolding or dissociating to aggregation-prone monomers, the reverse reaction of the aggregation-prone monomers folding or assembling back into the non-aggregating precursors may not be negligible. This process can be described as a reversible source-term, with a reverse step of reaction order n. For example, a folding/unfolding source-term has reaction order $n = 1$ and a tetramer

Table S2: Boundaries function between different regions.

Boundaries	Boundaries functions
1-3[1]	$\log_{10}\kappa \approx \log_{10} k_{source} + 1$
1-2	$\log_{10}\kappa \approx \frac{n_c - n_2}{n_c+3} \cdot \log_{10} k_{source} + \frac{n_2+3}{n_c+3} \cdot \log_{10}\lambda$
3-4	$\log_{10}\kappa \approx \log_{10}\lambda + 1$
2-4	$\log_{10}\lambda \approx \log_{10} k_{source} - 1$
1-5	$\log_{10}\kappa \approx \log_{10} k_{source} + 4$
2-5	$\log_{10}\lambda \approx \log_{10} k_{source} + 2.5$

dissociation/association source-term has reaction order $n = 4$. We consider reversible source terms of the form:

$$A \underset{k_{sink}}{\overset{k_{source}}{\rightleftharpoons}} n \cdot m \rightarrow \text{aggregation.} \tag{S33}$$

where A is the non-aggregating precursor and m is the aggregation-prone monomer. k_{source} is the rate constant of the forward source-term reaction which produces monomers whereas k_{sink} is the rate constant of its reverse. The differential equations describing this reversible source-term are:

$$\frac{dA(t)}{dt} = -k_{source} \cdot A(t) + k_{sink} \cdot m(t)^n \tag{S34a}$$

$$n \cdot A(t) + m(t) = m_{tot} - M(t). \tag{S34b}$$

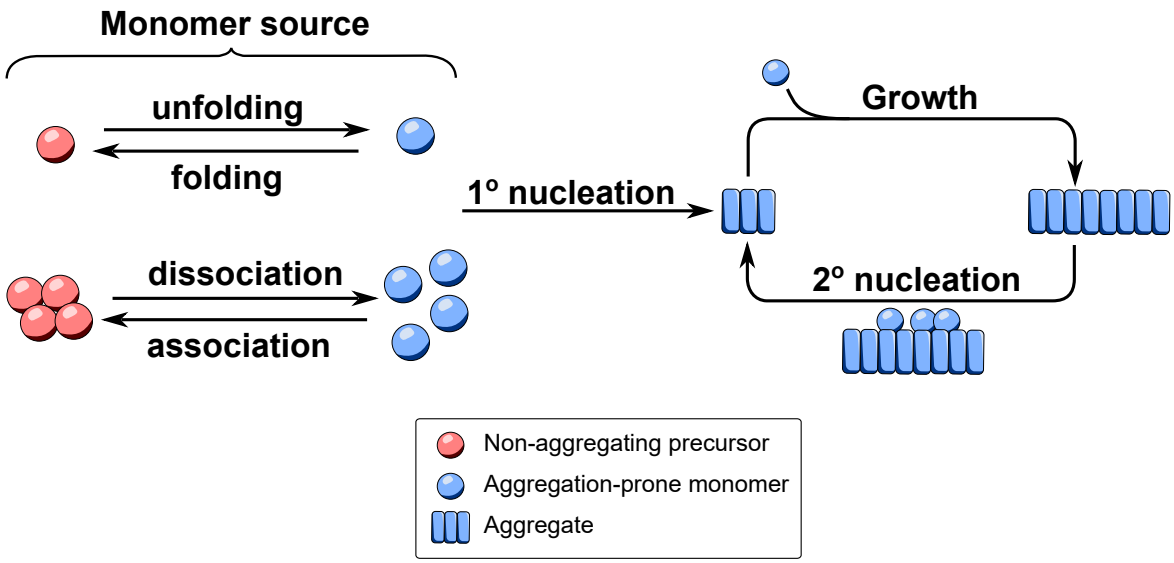

Figure S8: Elementary steps of protein aggregation with a reversible source-term. The reaction order of reverse reaction is $n = 1$ for an unfolding/folding source-term and $n = 4$ for a tetramer dissociation/association source-term.

where Eqs. (S34b) is the conservation of mass. m_{tot} is the total concentration of precursors, monomers and aggregates, in terms of monomer equivalents.

When deriving an analytical solution of aggregation with a reversible source-term, we consider two different regimes: the fast source regime (monomer production is much faster than aggregation: $\kappa \ll k_{source}, k_{sink}$) and the slow source regime (monomer production is much slower than aggregation: $\kappa \gg k_{source}, k_{sink}$).

In the fast source regime, we can assume that monomer production is in equilibrium, and thus ignore the differential equation for $A(t)$, Eqs. (S34a). The numerical plots of $m(t)$ and $n \cdot A(t)$ for different values of the total concentration $m_{tot} - M(t)$ are shown in Fig. S9.

For a general source-term of order n, Eqs. (S34) don't have an analytical solution. However, we can derive two extreme limits analytically: For $n = 1$, Eqs. (S34) can be

analytically solved:

$$A(t) = \frac{k_{sink}}{k_{sink} + k_{source}}(m_{tot} - M(t)) \tag{S35a}$$

$$m(t) = \frac{k_{source}}{k_{sink} + k_{source}}(m_{tot} - M(t)). \tag{S35b}$$

In these conditions the ratio of $A(t)$ and $m(t)$ is constant. By contrast, when $n \gg 1$, the systems behaviour begins to resemble a phase transition, with $m(t) = m_{tot} - M(t)$ for small $m_{tot} - M(t)$ and $m(t) = constant$ for large $m_{tot} - M(t)$. This regime would be encountered for example if $A(t)$ is an off-pathway condensate or amorphous precipitate.

Given these different regimes, we need to not only discuss whether the reaction has a fast source or slow source, but also consider if the source-term order n is small ($n = 1$) or large ($n = 10$). The numerical plots of aggregation in these four different regimes are shown in Fig. S10.

For $n = 1$ and a fast source-term. The concentrations of non-aggregating precursor $A(t)$ and the aggregation-prone monomers $m(t)$ remain proportional throughout. By applying the fast equilibrium approximation $A'(t) = 0$, the moment equations and the conservation law can be expressed as:

$$\frac{dP(t)}{dt} = k_n m(t)^{n_c} + k_2 m(t)^{n_2} M(t) \tag{S36a}$$

$$\frac{dM(t)}{dt} = 2k_+ m(t)^{n_+} P(t) \tag{S36b}$$

$$\frac{k_{source} + k_{sink}}{k_{source}} \cdot m(t) + M(t) = m_{tot} \tag{S36c}$$

$$M(0) = P(0) = 0 \tag{S36d}$$

By substituting the time-independent parameters

$$\begin{aligned} M(t)^{\text{eff}} &= M(t) \cdot \frac{k_{source}}{k_{source}+k_{sink}} \\ k_2^{\text{eff}} &= k_2 \cdot \frac{k_{source}+k_{sink}}{k_{source}} \\ k_+^{\text{eff}} &= k_+ \cdot \frac{k_{source}}{k_{source}+k_{sink}} \\ m_{tot}^{\text{eff}} &= m_{tot} \cdot \frac{k_{source}}{k_{source}+k_{sink}} \end{aligned}$$

into the integrated rate laws of aggregation without a source-term, Eqs. (S36) can be analytically solved.

This analytical solution for $M(t)$ is given in Eqs. (S37) and is plotted as the black dotted line in Fig. S10A, which fits well with the numerical solution (the cyan line).

$$M(t) = m_{tot} \cdot \left(1 - \left(1 + \frac{k_{source} k_n \left(\frac{k_{source} m_{tot}}{k_{sink}+k_{source}}\right)^{n_c - n_2 - 1} (2n_2+1) e^{\sqrt{2k_+ k_2 \left(\frac{k_{source} m_{tot}}{k_{sink}+k_{source}}\right)^{n_2+1}} \cdot t}}{6k_2(k_{sink}+k_{source})} \right)^{-\frac{3}{2n_2+1}} \right). \tag{S37}$$

For $n = 10$ and a fast source-term. The monomer concentration remains constant in the early time, until the non-aggregating precursor $A(t)$ is depleted.

The early time free monomer concentration is same as the monomer equilibrium concentration: $m_0 = m_{eq}$, which can be numerically solved for general n, as shown in Fig. S9B. The moment equations at the early time are then as follows:

$$P_0'(t) = k_n m_0^{n_c} + k_2 m_0^{n_2} \cdot M_0(t) \tag{S38a}$$

$$M_0'(t) = 2k_+ m_0^{n_+} \cdot P_0(t) \tag{S38b}$$

$$P_0(0) = M_0(0) = 0. \tag{S38c}$$

From which we can derive the early time solution $M_0(t)$. The early time solution $M_0(t)$ is:

$$M_0(t) = \frac{2k_n m_0^{n_c - n_2}}{k_2} \cdot Sinh^2[\frac{\sqrt{2k_+ k_2 m_0^{n_2+1}}}{2} \cdot t]. \tag{S39}$$

which agrees with the early time solution as obtained in the original derivation of the integrated rate laws,[14] but in this situation it is accurate for much longer, as the monomer concentration remains approximately constant until the precursor $A(t)$ is depleted. We can derive this critical time t_c at which the non-aggregating precursor $A(t)$ is depleted: $M_0(t_c) = m_{tot} - m_0$.

After $A(t)$ is depleted, the aggregation can be regarded simply as a seeded aggregation reaction with $m(t) = m_0$ and a start time of: t_c. The moment equations of this second stage are:

$$P'(\tau) = k_2 m(\tau)^{n_2} \cdot M_1(\tau) \tag{S40a}$$

$$M'(\tau) = 2k_+ m(\tau)^{n_+} \cdot P(\tau) \tag{S40b}$$

$$m(\tau) + M(\tau) = m_{tot} \tag{S40c}$$

$$M(0) = m_{tot} - m_0. \tag{S40d}$$

where $\tau = t - t_c$. This is simply a seeded reaction in the absence of a monomer source. We denote this solution here by $M_1(t)$.

The early and late time solutions match in both value and derivative at the critical time, i.e. $M_0(t_c) = M_1(t_c)$ and $M_0'(t_c) = M_1'(t_c)$. Thus we can directly unify the early time solution $M_0(t)$ and the late time solution $M_1(t)$ to express the full timescale solution $M(t)$, shown as the black dotted line in Fig. S10C.

We have discussed reversible source term with the initial condition $m(0) = 0$. Using

the same analytical approach, we can derive the approximate analytical solution for varies of initial conditions, such as $m(0) = m_{eq}$ and $m(0) = m_{tot}$.

For $n = 1$ and a slow source-term. Under the early time approximation (aggregated fibrils mass concentration is much smaller than the free monomer concentration) $M(t) \ll m(t)$, the reversible source-term ODE can be approximately written as:

$$\frac{dS_1(t)}{dt} = k_{source} \cdot (m_{tot} - S_1(t)) - k_{sink} \cdot S_1(t). \tag{S41}$$

where $S_1(t)$ is the source-term at early time and $S_1(t) = m(t) + M(t), S_1(0) = 0$. From Eqs. (S41), we obtain:

$$S_1(t) = \frac{k_{source}}{k_{source} + k_{sink}} \cdot m_{tot} \cdot [1 - e^{-(k_{source}+k_{sink})\cdot t}]. \tag{S42}$$

Based on the source term $S_1(t)$, we obtain the early time solution $M_0(t)$ (by substituting k_{source} with $k_{source}^{\text{eff}} = k_{source} + k_{sink}$ and substituting m_{tot} with $m_{tot}^{\text{eff}} = m_{tot} \cdot \frac{k_{source}}{k_{source}+k_{sink}}$ in Eqs. (5)).

Since the aggregation is much faster than monomer source and sink processes, aggregation quickly consumes all free monomers at the lag time. The newly produced monomers can be immediately consumed by aggregation, which implies the sink process can be neglected. The time evolution from this point on, $M_\infty(t)$, is simply given by the late time source-term function $S_2(t)$, which can be described by:

$$\frac{dS_2(t)}{dt} = k_{source} \cdot (m_{tot} - S_2(t)). \tag{S43}$$

The boundary condition of Eqs. (S43) is $S_2(t_{lag}) = S_1(t_{lag})$, where the lag time t_{lag} can be analytically derived from $M_0(t_{lag}) = 0.1 m_{tot}$. In such boundary condition, we derive

the late time source term $S_2(t)$ as:

$$S_2(t) = (m_{tot} - S_1(t_{lag})) \cdot (1 - e^{-k_{source} \cdot (t - t_{lag})}) + S_1(t_{lag}). \tag{S44}$$

which is also the late time solution $M_\infty(t) = S_2(t)$. Since the new produced monomers can quickly aggregate into fibrils, the exponential rate of $S_2(t)$ only depends on k_{source}. The full time solution of $M(t)$ can be expressed by unifying $M_0(t)$ and $M_\infty(t)$ using Eqs. (6). This analytical $M(t)$ is plotted as the black dotted line in Fig. S10B, which fits well with the numerical solution (the cyan line).

For $n = 10$ and a slow source-term. The method of deriving an integrated rate law is very similar to the $n = 1$ and slow source-term case. The reaction order is only relevant when the reverse reaction is significant, thus only affects the early time behaviour. Under the early time approximation $M(t) \ll m(t)$, the reversible source-term ODE can be approximately written as:

$$\frac{dS_{1,n}(t)}{dt} = k_{source} \cdot (m_{tot} - S_{1,n}(t)) - n \cdot k_{sink} \cdot S_{1,n}(t)^n. \tag{S45}$$

where $S_{1,n}(t)$ is the source-term at early time and $S_{1,n}(t) = m(t) + M(t)$ and the boundary condition is $S_{1,n}(0) = 0$. We can approximate the last term of Eqs. (S45) as m_{eq}^n and obtain:

$$S_{1,n}(t) = (m_{tot} - n \cdot \frac{k_{sink}}{k_{source}} \cdot m_{eq}^n) \cdot [1 - e^{-k_{source} \cdot t}]. \tag{S46}$$

where n is the order of the reversible source-term. Similarly as in the $n = 1$ case, we can write the late time source-term using Eqs. (S44).

The unified full time solution $M(t)$ has similar form as in the $n = 1$ case, by changing the source terms $S_1(t), S_2(t)$ to $S_{1,n}(t), S_{2,n}(t)$ (substitute $k_{source}^{\text{eff}} = k_{source}$ and $m_{tot}^{\text{eff}} =$

$m_{tot} - n \cdot \frac{k_{sink}}{k_{source}} \cdot m_{eq}^n$ into Eqs. (5)). This analytical $M(t)$ is shown as the black dotted line in Fig. S10D, which fits well with the numerical solution (the cyan line).

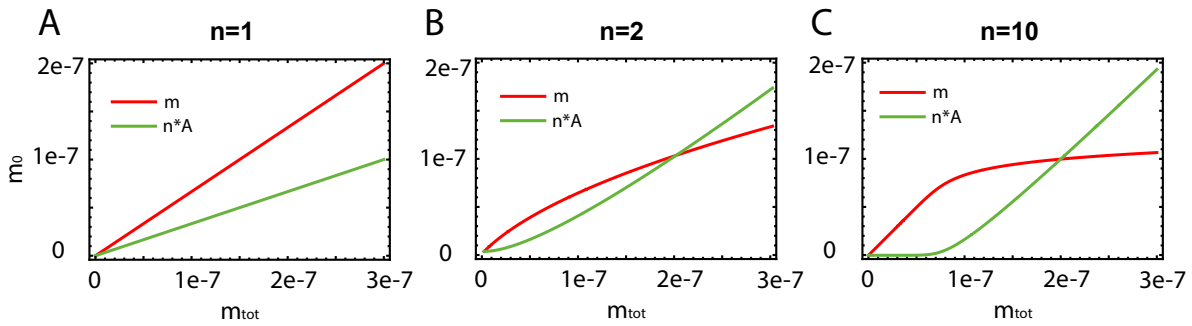

Figure S9: Numerical solution for the precursor (green) and monomer (red) concentrations at equilibrium as a function of the total protein concentration. The red curves are m and the green curves are $n \cdot A$. (A) is $n = 1$. (B) is $n = 2$. (C) is $n = 10$.

Besides the initial condition $m(0) = 0$, we also derived analytical solution of other initial conditions, such as: $m(0) = m_{eq}$ and $m(0) = m_{tot}$. The numerical and analytical plots for different initial conditions and $n = 1, 10$ are shown in Fig.S11.

For n=1, slow source-term with different initial conditions.

When the initial condition of Eqs. (S41) is $S_1(0) = m_0$. The early time source-term can be derived as:

$$S_1(t) = (\frac{k_{source} \cdot m_{tot}}{k_{source} + k_{sink}} - m_0) \cdot [1 - e^{-(k_{source}+k_{sink})\cdot t}] + m_0. \tag{S47}$$

For the late time, fast aggregation can immediately consumes all newly produced monomers, which implies the monomer sink process can be neglected. We use the same ODE Eqs. (S43) to derive the late time source-term under the boundary condition $S_2(t_{lag}) =$

$S_1(t_{lag})$:

$$M_\infty(t) = S_2(t) = (m_{tot} - S_1(t_{lag})) \cdot (1 - e^{-k_{source} \cdot (t - t_{lag})}) + S_1(t_{lag}) \tag{S48}$$

where the lag time t_{lag} can be analytically derived from $M_0(t_{lag}) = 0.1 m_{tot}$.

For initial conditions $m(0) = 0, m_{eq}, m_{tot}$, This analytical full time solution $M(t)$ is shown as the black dotted line in Fig. S11A-C, which fits well with the numerical solution (the cyan line).

For n=10, slow source-term with different initial conditions.

When the initial condition of Eqs. (S45) is $S_{1,n}(0) = m_0$ and we similarly approximate the $S_{1,n}(t)$ of its last term as m_{eq}. The early time source-term can be derived as:

$$S_{1,n}(t) = (m_{tot} - n \cdot m_{eq}^n \cdot \frac{k_{sink}}{k_{source}}) \cdot (1 - e^{-k_{source} \cdot t}) + m_0 e^{-k_{source} \cdot t}. \tag{S49}$$

For the late time, fast aggregation can immediately consumes all newly produced monomers, which implies the monomer sink process can be neglected. We use the same ODE Eqs. (S48) to derive the late time source-term under the boundary condition $S_{2,n}(t_{lag}) = S_{1,n}(t_{lag})$.

For initial conditions $m(0) = 0, m_{eq}, m_{tot}$, This analytical full time solution $M(t)$ is shown as the black dotted line in Fig. S11D-F, which fits well with the numerical solution (the cyan line).

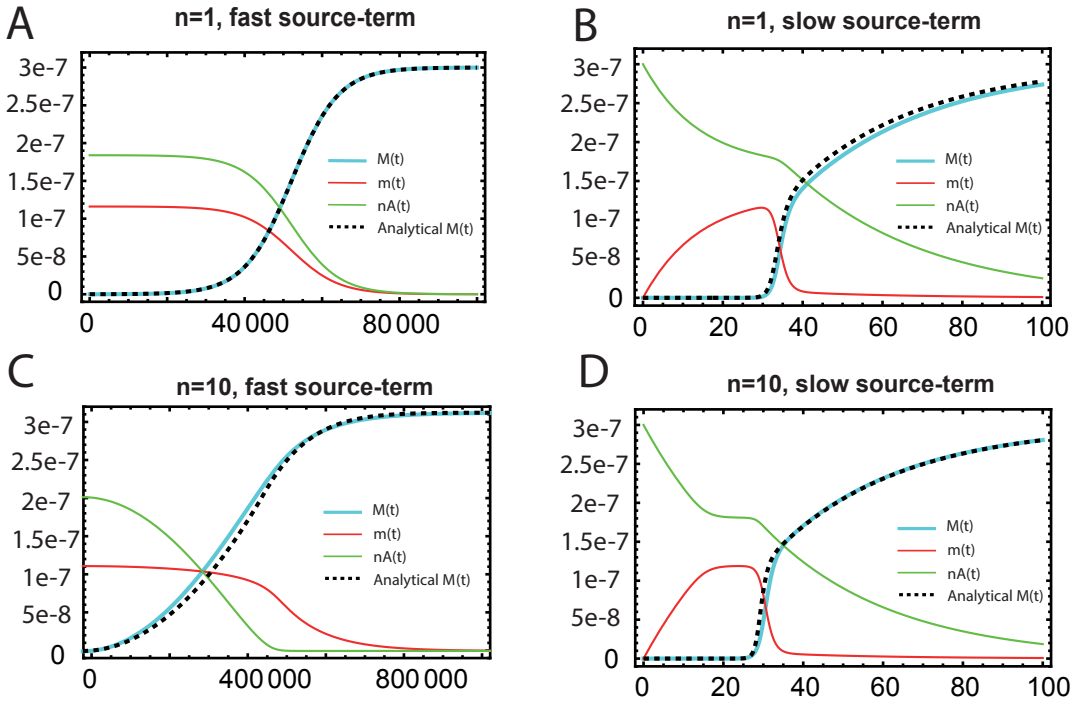

Figure S10: Numerical and analytical integrated rate laws of aggregation with a reversible source-term with the initial condition $m(0) = 0$**.** (A) shows a fast source-term (monomer production rate constants are much larger than the aggregation rate) $k_{source}, k_{sink} \gg \kappa$ with a low reaction order $n = 1$. (B) shows a slow source-term (monomer production rate constants are much smaller than the aggregation rate) $k_{source}, k_{sink} \ll \kappa$ with a low reaction order $n = 1$. (C) shows a fast source-term $k_{source}, k_{sink} \gg \kappa$ with a high reaction order $n = 10$. (D) shows a slow source-term $k_{source}, k_{sink} \ll \kappa$ with a high reaction order $n = 10$. Cyan curves are $M(t)$ obtained from numerically integrating Eqs. (S34) and the moment equations Eqs. (1). Red curves are the numerical $m(t)$. Green curves are the numerical $n \cdot A(t)$. The black dashed curves are the analytical solutions.

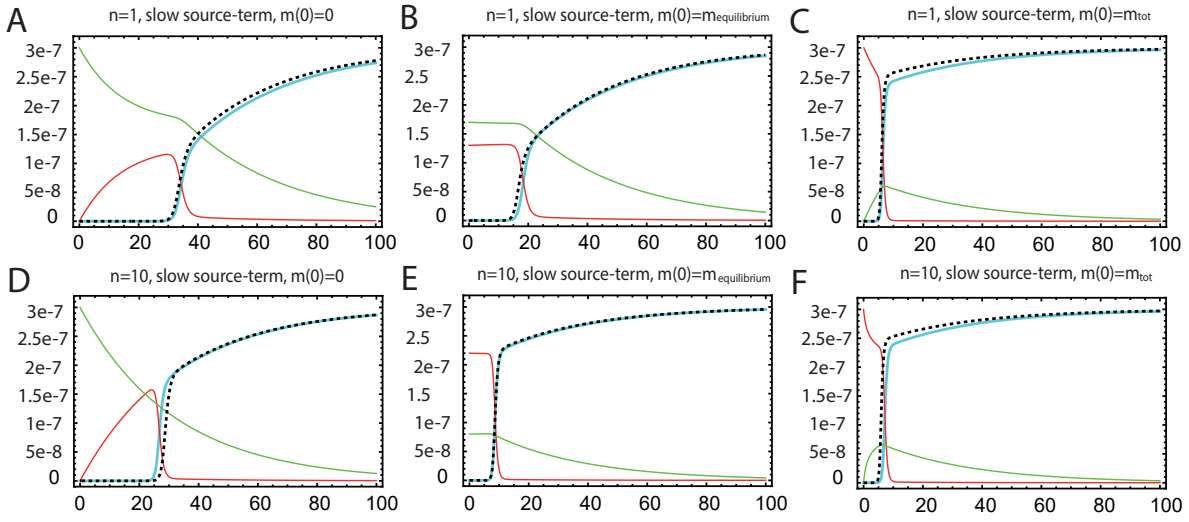

Figure S11: Numerical and analytical integrated rate laws of aggregation with a reversible source-term with the initial conditions: $m(0) = 0$ for (A) and (D), $m(0) = m_{eq}$ for (B) and (E) and $m(0) = m_{tot}$ for (C) and (F). The reversible source-term orders: $n = 1$ for (A)-(C) and $n = 10$ for (D)-(F). Red curves are the numerical $m(t)$. Green curves are the numerical $n \cdot A(t)$. The black dashed curves are the analytical solutions.

References

(1) Månsson, C.; Kakkar, V.; Monsellier, E.; Sourigues, Y.; Härmark, J.; Kampinga, H. H.; Melki, R.; Emanuelsson, C. DNAJB6 is a peptide-binding chaperone which can suppress amyloid fibrillation of polyglutamine peptides at substoichiometric molar ratios. *Cell Stress and Chaperones* **2013**, *19*, 227–239.

(2) Meisl, G.; Kirkegaard, J. B.; Arosio, P.; Michaels, T. C. T.; Vendruscolo, M.; Dobson, C. M.; Linse, S.; Knowles, T. P. J. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nature Protocols* **2016**, *11*, 252–272.

(3) Knowles, T. P. J.; Buehler, M. J. Nanomechanics of functional and pathological amyloid materials. *Nature Nanotechnology* **2011**, *6*, 469–479.

(4) Arosio, P.; Cedervall, T.; Knowles, T. P.; Linse, S. Analysis of the length distribution of amyloid fibrils by centrifugal sedimentation. *Analytical Biochemistry* **2016**, *504*, 7–13.

(5) Nasir, I.; Linse, S.; Cabaleiro-Lago, C. Fluorescent Filter-Trap Assay for Amyloid Fibril Formation Kinetics in Complex Solutions. *ACS Chemical Neuroscience* **2015**, *6*, 1436–1444.

(6) Gaspar, R.; Meisl, G.; Buell, A. K.; Young, L.; Kaminski, C. F.; Knowles, T. P. J.; Sparr, E.; Linse, S. Secondary nucleation of monomers on fibril surface dominates α-synuclein aggregation and provides autocatalytic amyloid amplification. *Quarterly Reviews of Biophysics* **2017**, *50*, missing.

(7) Knowles, T. P. J.; Waudby, C. A.; Devlin, G. L.; Cohen, S. I. A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. An Analytical

Solution to the Kinetics of Breakable Filament Assembly. *Science* **2009**, *326*, 1533–1537.

(8) Cohen, S. I. A.; Vendruscolo, M.; Welland, M. E.; Dobson, C. M.; Terentjev, E. M.; Knowles, T. P. J. Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments. *The Journal of Chemical Physics* **2011**, *135*, 065105.

(9) Cohen, S. I. A.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Nucleated polymerization with secondary pathways. II. Determination of self-consistent solutions to growth processes described by non-linear master equations. *The Journal of Chemical Physics* **2011**, *135*, 065106.

(10) Cohen, S. I. A.; Linse, S.; Luheshi, L. M.; Hellstrand, E.; White, D. A.; Rajah, L.; Otzen, D. E.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Proliferation of amyloid-β 42 aggregates occurs through a secondary nucleation mechanism. *Proceedings of the National Academy of Sciences* **2013**, *110*, 9758–9763.

(11) Meisl, G.; Yang, X.; Hellstrand, E.; Frohm, B.; Kirkegaard, J. B.; Cohen, S. I. A.; Dobson, C. M.; Linse, S.; Knowles, T. P. J. Differences in nucleation behavior underlie the contrasting aggregation kinetics of the $A\beta40$ and $A\beta42$ peptides. *Proceedings of the National Academy of Sciences* **2014**, *111*, 9384–9389.

(12) Hartman, P. *Ordinary Differential Equations*; Society for Industrial and Applied Mathematics, 2002.

(13) Dear, A. J.; Meisl, G.; Linse, S.; Mahadevan, L. Developing integrated rate laws of complex self-assembly reactions using Lie symmetry: Kinetics of Abeta42, Abeta40 and Abeta38 co-aggregation. *arXiv* **2023**, *2309.05038*, missing.

(14) Knowles, T. P. J.; White, D. A.; Abate, A. R.; Agresti, J. J.; Cohen, S. I. A.; Sperling, R. A.; Genst, E. J. D.; Dobson, C. M.; Weitz, D. A. Observation of spatial propagation of amyloid assembly from single nuclei. *Proceedings of the National Academy of Sciences* **2011**, *108*, 14746–14751.