AIP | The Journal of Chemical Physics

# The length distribution of frangible biofilaments

Thomas C. T. Michaels, Pernille Yde, Julian C. W. Willis, Mogens H. Jensen, Daniel Otzen, Christopher M. Dobson, Alexander K. Buell, and Tuomas P. J. Knowles

Citation: The Journal of Chemical Physics **143**, 164901 (2015); doi: 10.1063/1.4933230
View online: http://dx.doi.org/10.1063/1.4933230
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/143/16?ver=pdfcov
Published by the AIP Publishing



# The length distribution of frangible biofilaments

Thomas C. T. Michaels,[1] Pernille Yde,[2] Julian C. W. Willis,[1] Mogens H. Jensen,[2] Daniel Otzen,[3] Christopher M. Dobson,[1] Alexander K. Buell,[1] and Tuomas P. J. Knowles[1,a)]
[1]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*
[2]*Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark*
[3]*Interdisciplinary Nanoscience Center, Department of Molecular Biology and Genetics, Center for Insoluble Protein Structures, Aarhus University, Gustav Wieds Vej 14, 8000 Aarhus C, Denmark*

(Received 27 April 2015; accepted 4 October 2015; published online 22 October 2015)

A number of different proteins possess the ability to polymerize into filamentous structures. Certain classes of such assemblies can have key functional roles in the cell, such as providing the structural basis for the cytoskeleton in the case of actin and tubulin, while others are implicated in the development of many pathological conditions, including Alzheimer's and Parkinson's diseases. In general, the fragmentation of such structures changes the total number of filament ends, which act as growth sites, and hence is a key feature of the dynamics of filamentous growth phenomena. In this paper, we present an analytical study of the master equation of breakable filament assembly and derive closed-form expressions for the time evolution of the filament length distribution for both open and closed systems with infinite and finite monomer supply, respectively. We use this theoretical framework to analyse experimental data for length distributions of insulin amyloid fibrils and show that our theory allows insights into the microscopic mechanisms of biofilament assembly to be obtained beyond those available from the conventional analysis of filament mass only. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4933230]

## I. INTRODUCTION

The spontaneous formation of filamentous protein structures from soluble monomers is a process of fundamental importance to the normal functioning of biological systems[1–6] but is also often encountered as the mechanism responsible for the formation of protein deposits observed in association with many neurodegenerative disorders, including Parkinson's, Alzheimer's, and prion diseases.[7,8] As such, numerous experimental and theoretical studies in the literature have focussed on understanding the physical principles underlying the kinetics of the self-assembly of filamentous protein structures. In addition to the classical picture of primary nucleation events followed by sequential elongation steps,[1,2] fragmentation has been shown to be a crucial feature characterising the kinetics of many filamentous assembly systems.[9–17] Indeed, the breakage of long filaments into shorter ones increases the number of free ends, which act as growth sites, and hence accelerates the overall growth reaction. An important result of recent biophysical research is the progress towards the description of the time evolution of the principal moments of the filament distribution;[16–22] these are average quantities which relate to common experimental observables, including the total number and mass concentrations of aggregates or the average length of filaments, and obey a closed set of few coupled differential equations, which can be treated analytically using self-consistent methods,[16,17,23] providing a route for connecting macroscopic measurements of filamentous protein kinetics with the underlying microscopic processes. Yet, in many cases, knowledge of the detailed nature of the length distribution is desirable, for instance, because smallest aggregate species have been associated with higher cellular toxicity in the context of amyloid formation.[24–26] In this context, previous work has focussed on the steady-state properties of the filament length distribution,[22,27] but in the absence of closed-form expressions for the time evolution of the aggregate size distribution, many studies have instead relied on numerical methods for integrating the kinetic equations.[12,14,27–29]

Building on previous results,[21,22,27] we present here closed-form expressions for the time evolution of the mass distribution of breakable filamentous structures. We show that our theoretical framework allows effective exploration of the parameter space and identifies the key physical factors that govern the distribution of aggregate length in breakable filament assembly. Furthermore, we use the resulting analytical expressions to fit experimental length distributions obtained from measurements of growing insulin amyloid fibrils.

## II. MASTER EQUATION

We consider a large ensemble of monomeric polypeptide molecules and aggregates that undergo the growth processes outlined in Fig. 1: fibrils are formed initially through primary nucleation and subsequently increase in size through linear growth, i.e., through the addition/removal of monomeric molecules onto/from the ends of the filament; in addition, aggregates are able to multiply in number through breakage. In the mean-field limit, the time evolution of the concentrations of the resulting aggregate structures obeys the law of mass action in

a)tpjk2@cam.ac.uk

 

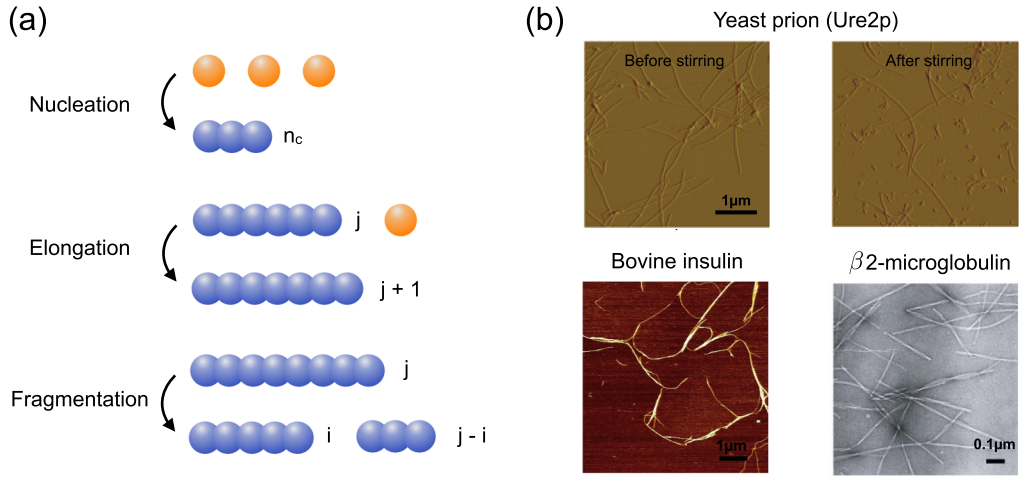


FIG. 1. (a) Schematic representation of the elementary mechanistic steps of breakable filament assembly. (b) Representative examples of filamentous protein systems exhibiting fragmentation: AFM/TEM images of fibrillization of yeast prion (Ure2p),[10] $\beta$2-microglobulin,[11] and bovine insulin. Part (b) reproduced with permission from Tanaka *et al.*, Nature **442**, 585 (2006), copyright 2006 Macmillan Publisher Ltd. and Xue *et al.*, Proc. Natl. Acad. Sci. U. S. A. **105**, 8926 (2008), copyright 2008 National Academy of Sciences, USA.

form of a mass balance equation,[16–22,30]

$$\begin{aligned}\frac{\partial f(t,j)}{\partial t} &= 2k_+m(t)f(t,j-1) - 2k_+m(t)f(t,j)\\ &\quad + 2k_{\rm off}f(t,j+1) - 2k_{\rm off}f(t,j)\\ &\quad - k_-(j-1)f(t,j) + 2k_-\sum_{i=j+1}^{\infty} f(t,i)\\ &\quad + k_n m(t)^{n_c}\delta_{j,n_c},\end{aligned} \tag{1}$$

where $f(t,j)$ denotes the concentration of filaments of length $j$ at time $t$ and $k_+$, $k_{\rm off}$, $k_-$ and $k_n$ are the rate constants for filament elongation, monomer dissociation, fibril breakage, and homogeneous (primary) nucleation, respectively. The terms in Eq. (1) proportional to $2k_+m(t)$ describe the reactive flux associated with aggregate elongation, whereby the factor 2 accounts for the fact that each filament has two active ends. Similarly, terms proportional to $2k_{\rm off}$ express the dissociation of monomers from filament ends. The third line of Eq. (1) pertains to aggregate fragmentation; the term $k_-(j-1)f(t,j)$ describes the loss of polymers of length $j$ when they break into smaller fragments, whereas the sum $2k_-\sum_{i>j} f(t,i)$ accounts for the creation of filaments of size $j$ as a result of the breakage of longer ones. The last line of Eq. (1) describes the generation of growth-competent aggregates through primary nucleation, with $n_c$ giving the overall dependency of this processes on the concentration of free monomers, which might, in general, involve more than one elementary step.[18,31,32] The condition $f(t,j) = 0$ is imposed for $j < n_c$. Note that in our formalism, breakage implicitly contributes to the dissociation of polymers. Monomers are, therefore, removed from the ends of filaments with the effective rate constant $k_{\rm off} + k_-$. Moreover, in Eq. (1), we have assumed size-independent rate constants. This assumption was primarily motivated by the fact that current experimental data do not provide reliable estimates for the size dependence of the rates. Our theoretical framework, however, can in principle be extended to take this effect into account.[33]

## III. BREAKABLE FILAMENT ASSEMBLY IN OPEN SYSTEMS

We first examine the time evolution of the full aggregate size distribution in an open system, where the concentration of monomers remains constant at the initial value $m(t) = m(0)$, for instance, through the action of protein synthesis in a living cell. A discussion on mass-conserving systems is given in Sec. IV. Under these circumstances, the underlying master equation, Eq. (1), can be solved exactly to give the full aggregate size distribution in closed form (see Appendix A for details); however, the resulting expressions are very complicated. In order to simplify the functional form of the resulting length distribution, we consider here a continuum limit approximation of Eq. (1) and describe the aggregate size $j$ as being a continuous variable $x$. The quantity of interest is therefore $f(t,x)dx$, the concentration of aggregates with size in the range $(x, x+dx)$ at time $t$. Within an open system scenario, this quantity evolves according to the following master equation in the continuum-limit (see Appendix B for a derivation of this result):[22,27,34]

$$\begin{aligned}\frac{\partial f(\tau,x)}{\partial \tau} &= -\xi\frac{\partial f(\tau,x)}{\partial x} - xf(\tau,x) + 2\int_x^{\infty} f(\tau,z)dz\\ &\quad + \nu\delta(x-n_c),\end{aligned} \tag{2}$$

where, for convenience, we have introduced the rescaled time variable,

$$\tau = k_-t \tag{3}$$

and the dimensionless parameters

$$\xi = \frac{2[k_+m(0) - k_{\rm off}]}{k_-}, \quad \nu = \frac{k_n m(0)^{n_c}}{k_-} \tag{4}$$

measuring the driving forces of linear growth and primary nucleation relative to fragmentation. Note that for $k_{\rm off} = 0$, such that the rate constants for monomer dissociation and fibril fragmentation are the same, the parameter $\xi$ is related to the familiar saturation concentration.

For $x > n_c$, the nucleation term in Eq. (2) vanishes and we can solve Eq. (2) using the Laplace transform $\hat{f}(s,x)$

$= \int_0^\infty f(\tau,x)e^{-s\tau}d\tau$. In terms of this transform, Eq. (2) becomes after differentiation with respect to $x$,

$$\xi\frac{\partial^2\hat{f}(s,x)}{\partial x^2} + (x+s)\frac{\partial\hat{f}(s,x)}{\partial x} + 3\hat{f}(s,x) = 0, \tag{5}$$

where we have assumed that no seed material is present at the beginning of the reaction. The general solution of Eq. (5) is

$$\hat{f}(s,x) = A(s)e^{-\frac{x(x+2s)}{2\xi}}[(x+s)^2-\xi] + B(s)\frac{(x+s)\sqrt{\xi}+\sqrt{2}[\xi-(x+s)^2]D\left(\frac{x+s}{\sqrt{2\xi}}\right)}{\sqrt{\xi}}, \tag{6}$$

where $D(x) = e^{-x^2}\int_0^x e^{y^2}dy$ denotes the Dawson integral and $A(s)$ and $B(s)$ are ($s$-dependent) constants of integration determined by the boundary conditions. Because the integral over $(n_c,\infty)$ of the second term in Eq. (6) is divergent,[35] resulting in an infinite polymer number concentration, we must have $B(s) = 0$. In order to fix the value of $A(s)$, we note that the delta function at $x = n_c$ can be reformulated as boundary condition for $f(\tau,x)$ at $x = n_c$,

$$f(\tau,x=n_c) = \frac{\nu}{\xi} \tag{7}$$

or, in terms of the Laplace transform $\hat{f}(s,x)$,

$$\hat{f}(s,x=n_c) = \frac{\nu}{\xi s}. \tag{8}$$

Hence, by combining Eq. (6) with Eq. (8), we obtain

$$\hat{f}(s,x) = \frac{\nu e^{-\frac{(x-n_c)(x+n_c+2s)}{2\xi}}[(x+s)^2-\xi]}{\xi s[(n_c+s)^2-\xi]}. \tag{9}$$

The inverse Laplace transform of Eq. (9) is given by the inversion formula

$$f(\tau,x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\hat{f}(s,x)e^{s\tau}ds = \sum_{\text{Poles of }\hat{f}(s,x)e^{\tau s}}\text{Res}\,\hat{f}(s,x)e^{\tau s}, \tag{10}$$

where the contour of integration is the closure of any vertical line $c$ such that $\hat{f}(s,x)$ has no poles on or to the right of it. By the residue theorem, we can compute the inverse Laplace transform of $\hat{f}(s,x)$ as the sum of the residues of $\hat{f}(s,x)e^{\tau s}$ at poles of $\hat{f}(s,x)$. Equation (9) has simple poles at 0 and $\pm\sqrt{\xi}-n_c$, which correspond to constant and exponentially decaying/growing terms in $f(\tau,x)$, respectively. Therefore, after employing that for most systems of interest $\sqrt{\xi} \gg n_c$[36] and transforming back to real time $t = \tau/k_-$, the resulting expression for $f(t,x)$ obtained using Eq. (10) reads

$$f(t,x) = \frac{\nu e^{-\frac{x^2-n_c^2}{2\xi}}\left[\xi - x^2 + x^2\cosh\left(\kappa t + \frac{n_c-x}{\sqrt{\xi}}\right) + 2x\sqrt{\xi}\sinh\left(\kappa t + \frac{n_c-x}{\sqrt{\xi}}\right)\right]}{\xi^2}H(vt-x+n_c), \tag{11}$$

where $v = 2[k_+m(0) - k_{\text{off}}]$, $\kappa = \sqrt{2k_-[k_+m(0)-k_{\text{off}}]}$ is an effective rate constant for aggregate proliferation, and $H(x)$ is the Heaviside step function, defined by $H(x) = 1$ for $x > 0$ and $H(x) = 0$ for $x < 0$.

Equation (11) gives, in closed form, the temporal evolution of the length distribution of breakable filaments in a system where the concentration of monomers is constant in time. A comparison of the $f(t,x)$ profiles predicted by Eq. (11) and the numerical solution of the master equation[37] is shown in Fig. 2. At early times, filament elongation dominates over fragmentation and the length distribution behaves similarly to a shock wave moving in size space. The velocity of the wave front is $v = 2[k_+m(0) - k_{\text{off}}]$ and its position at time $t$ is $x_f = n_c + vt$. At later times, $\kappa t \gg 1$, the shock-wave nature of the length distribution disappears and the system approaches a stationary state, as a result of the dominance of the exponential growing term $e^{\kappa t}$,[22,27,38]

$$\frac{f(t,x)}{P(t)} \overset{\kappa t\gg 1}{\to} \frac{e^{-\frac{(x-n_c)(2\sqrt{\xi}+x+n_c)}{2\xi}}x(x+2\sqrt{\xi})}{\xi^{3/2}}, \tag{12}$$

where $P(t) \sim \nu/(2\sqrt{\xi})e^{\kappa t}$ is the polymer number concentration (see Eq. (15)). Under these circumstances, both $f(t,x)$ and the total number concentration of aggregates are exponentially growing with time with multiplication rate $\kappa$; hence, the fractional occupation $f(t,x)/P(t)$ stays constant in time and takes the form of a biased Gaussian, which is solely determined by the parameter $\xi$.

It is interesting to consider the limit of vanishing fragmentation rate; in this situation, Eq. (11) recovers the solution of the Oosawa model with constant monomer concentration[30]

$$f(t,x) \overset{k_-\to 0}{\to} \frac{k_n m(0)^{n_c-1}}{2k_+}H(vt-x+n_c), \tag{13}$$

where we set $k_{\text{off}} = 0$. In this limit, the master equation reduces to an advection equation; the aggregate mass distribution is therefore given by a shock wave moving in $x$ space with constant velocity $v$.

An interesting scenario is the inclusion in the master equation (1) of a term responsible for the degradation of polymers.[21] While this process does not play an important role in experiments of filamentous assembly *in vitro* that currently account for the bulk of the accurate available data, application to *in vivo* situations requires the inclusion of fibril degradation. Our framework can be extended to take this effect into account; an expression for the aggregate size distribution in the presence of polymer degradation is presented in Appendix C.

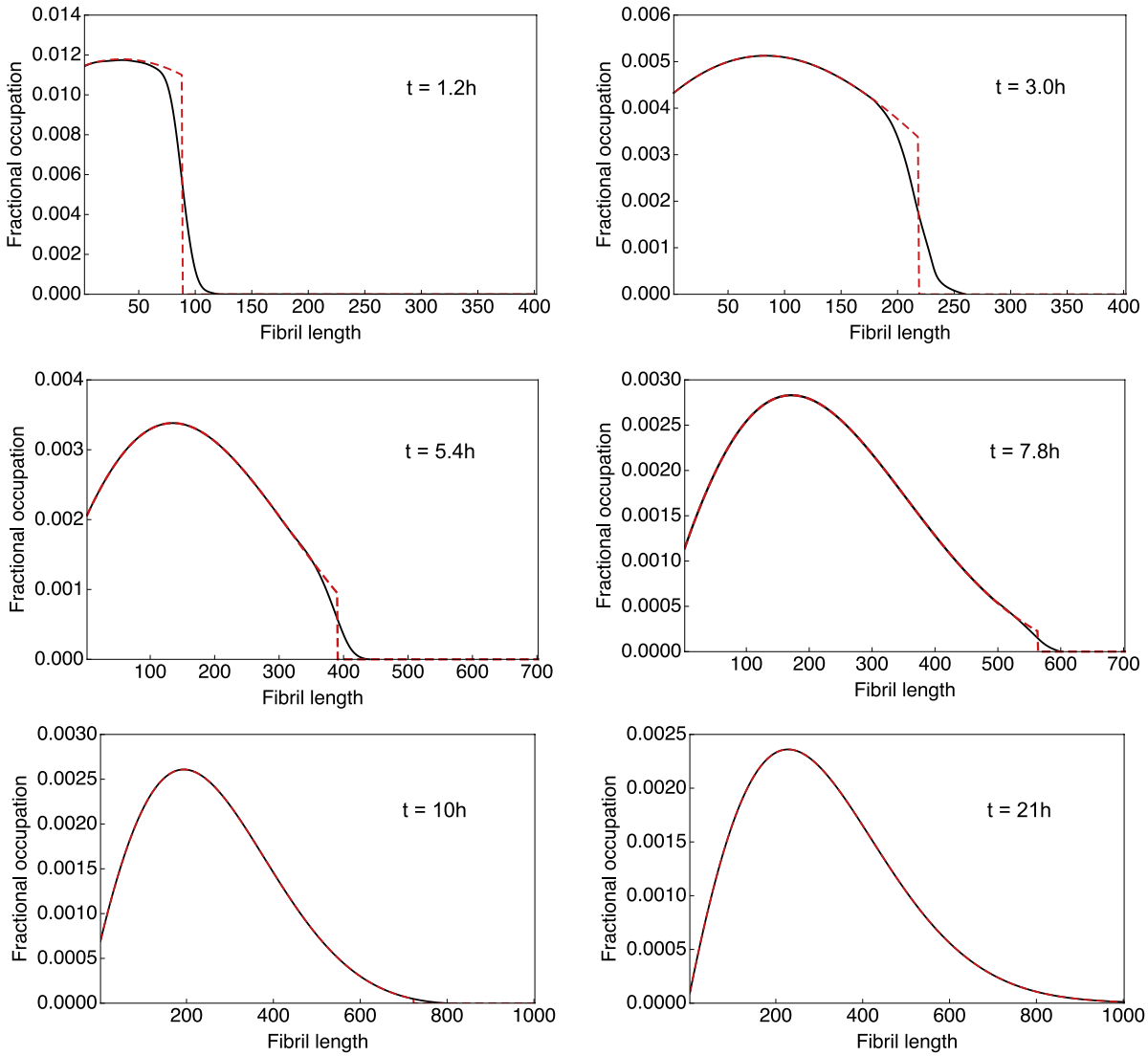


FIG. 2. Time evolution of fractional occupation $f(t,x)/P(t)$ of frangible filaments in an open system. The continuum-limit solution Eq. (11) (red dashed) is compared with the numerical solution of the master equation (black solid). The parameters are $k_+ = 10^4\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_{\mathrm{off}} = 0$, $k_n = 10^{-5}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_- = 2\times10^{-7}\ \mathrm{s}^{-1}$, $m(0) = 1\ \mu\mathrm{M}$, $n_c = 2$, and $M(0) = P(0) = 0$.

Finally, as a consistency check, we derive the time evolution of the first principal moments starting from the obtained expressions for the filament size distribution. For example, the Laplace transform of the polymer number concentration, $\hat{P}(s) = \int_0^\infty P(\tau)e^{-s\tau}d\tau$, is obtained through integration of Eq. (9) as

$$\hat{P}(s) = \int_{n_c}^{\infty} \hat{f}(s,x)dx = \frac{\nu(n_c+s)}{s(s-\sqrt{\xi})(s+\sqrt{\xi})}, \qquad (14)$$

where we used $\sqrt{\xi} \gg n_c$. Computing the inverse Laplace transform of Eq. (14) using Eq. (10) yields

$$P(t) = \frac{\nu}{2\sqrt{\xi}}\left(e^{\kappa t} - e^{-\kappa t}\right) - \frac{n_c\nu}{\xi} \qquad (15)$$

in agreement with previous reports.[16–20] Similarly, the following formula for the Laplace transform of the polymer mass concentration is obtained from Eq. (9):

$$\hat{M}(s) = \int_{n_c}^{\infty} x\hat{f}(s,x)dx = \frac{\nu[n_c(n_c+s)+\xi]}{s(s-\sqrt{\xi})(s+\sqrt{\xi})}, \quad \sqrt{\xi} \gg n_c. \qquad (16)$$

Computing the inverse Laplace transform of Eq. (16) using Eq. (10) yields therefore[16–20]

$$M(t) = \frac{\nu}{2}\left(e^{\kappa t} + e^{-\kappa t}\right) - \nu. \qquad (17)$$

Note that while both $P(t)$ and $M(t)$ continue to grow exponentially after the stationary distribution Eq. (12) is attained, the average length of fibrils, $L(t) = M(t)/P(t) \sim \sqrt{\xi}$, stays constant in this limit.[21]

## IV. BREAKABLE FILAMENT ASSEMBLY IN CLOSED SYSTEMS

Having elucidated the physical principles underlying breakable filament growth under constant monomer input, we now focus on the temporal evolution of the length distribution of filaments for mass-conserving systems, which are characterized by finite monomer supply.

Through numerical integration of master equation (1), it has been shown that the full time evolution of the length

distribution of breakable filaments in a mass-conserving system evolves through two key stages of dynamics:[27] during the early stages of the polymerization reaction, the concentration of monomers is approximatively constant at the value $m(0)$ and the aggregate size distribution evolves according to the solution Eq. (11) for an open system discussed in Sec. III; during the later stages of the reaction, when monomers have been depleted, the system enters a slow phase of dynamics where the length distribution shifts smoothly from stationary distribution (12) into a steady-state distribution dominated by small filament fragments. This end point of the assembly reaction is characterized by the following expressions for principal moments:[21,22,27]

$$M(\infty) = m(0) - \frac{k_- n_c(n_c-1)}{2k_+}, \quad P(\infty) = \frac{M(\infty)}{2n_c - 1}. \tag{18}$$

Relying on the observations above, we account for finite monomer supply, by assuming the following form for the time-varying concentration of monomers (Fig. 3):

$$m(t) = \begin{cases} m(0), & \text{for } t < T \\ m(\infty) = m(0) - M(\infty), & \text{for } t \geq T \end{cases}, \tag{19}$$

where[42]

$$T = \frac{1}{\kappa} \log\left(\frac{2M(\infty)}{\nu}\right) \tag{20}$$

is the time at which the expression Eq. (17) reaches steady-state, $M(\infty)$,

$$M(t) = \begin{cases} \frac{\nu}{2}\left(e^{\kappa t} + e^{-\kappa t}\right) - \nu, & \text{for } t < T \\ M(\infty), & \text{for } t \geq T \end{cases}. \tag{21}$$

Accordingly, the number concentration of aggregates is assumed to evolve as

$$P(t) = \begin{cases} \frac{\nu}{2\sqrt{\xi}}\left(e^{\kappa t} - e^{-\kappa t}\right) - \frac{n_c \nu}{\xi}, & \text{for } t < T \\ P(\infty) + \left(\frac{M(\infty)}{\sqrt{\xi}} - P(\infty)\right) e^{-(2n_c-1)k_- t}, & \text{for } t \geq T \end{cases}. \tag{22}$$

On assuming Eq. (19), the species distribution $f(t,j)$ evolves through two stages of dynamics.

### A. First time scale: $t < T$ (advection phase)

During this phase of the reaction, the length distribution develops according to Eq. (11).

### B. Second time scale: $t > T$ (redistribution phase)

At the end of the previous time scale, the aggregate size distribution has reached the stationary value

$$\frac{M(\infty) e^{-\frac{(j-n_c)(2\sqrt{\xi}+j+n_c)}{2\xi}} j(j+2\sqrt{\xi})}{\xi^2} =: f_{ss}(j). \tag{23}$$

To describe how Eq. (23) develops into the steady-state distribution, we introduce the rescaled time variable

$$\tau = k_-(t - T) \tag{24}$$

and solve Eq. (1) with $m(t) = m(\infty)$ and $k_n = 0$,

$$\frac{\partial f(\tau,j)}{\partial \tau} = n_c(n_c-1)f(\tau, j-1) - n_c(n_c-1)f(\tau,j) - (j-1)f(\tau,j) + 2\sum_{i=j+1}^{\infty} f(\tau,i) \tag{25}$$

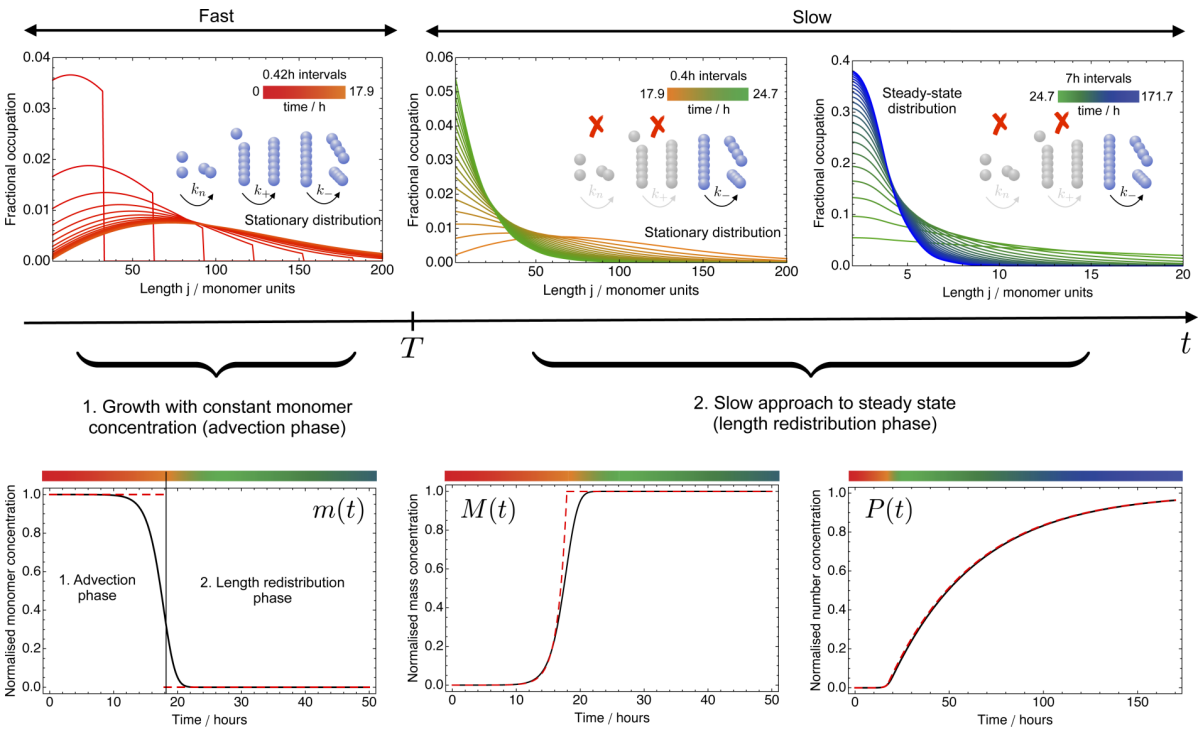


FIG. 3. The time evolution of the length distribution of breakable protein filaments occurs through two distinct phases: a fast advection phase followed by a slow length redistribution phase ending into a steady-state distribution dominated by short fragments. The curves were generated using Eqs. (11) and (34) for the following parameters: $k_+ = 10^4\ \mathrm{M^{-1}\ s^{-1}}$, $k_{\mathrm{off}} = 0$, $k_n = 10^{-5}\ \mathrm{M^{-1}\ s^{-1}}$, $k_- = 2\times10^{-6}\ \mathrm{s^{-1}}$, $n_c = 2$, $m(0) = 1\ \mu\mathrm{M}$, $P(0) = M(0) = 0$.

subject to the initial condition

$$f(\tau = 0, j) = f_{ss}(j). \tag{26}$$

Introducing the generating function[39–41]

$$C(z,\tau) = \sum_{j=n_c}^{\infty} f(\tau, j) z^j \tag{27}$$

and using the following results:

$$\sum_{j=n_c}^{\infty} j z^j f(\tau, j) = z \frac{\partial C(z,\tau)}{\partial z} \tag{28}$$

$$\sum_{j=n_c}^{\infty} z^j \sum_{i=j+1}^{\infty} f(\tau, i) = \sum_{j=n_c}^{\infty} f(\tau, j) \sum_{i=n_c}^{j-1} z^i = \frac{2 z^{n_c} P(\tau)}{1-z} - \frac{2C(z,\tau)}{1-z} \tag{29}$$

allow recasting Eq. (25) into the following partial differential equation:

$$\frac{\partial C(z,\tau)}{\partial \tau} + z \frac{\partial C(z,\tau)}{\partial z} = \left[ n_c(n_c - 1)(z-1) + 1 - \frac{2}{1-z} \right] C(z,\tau) + \frac{2 z^{n_c} P(\tau)}{1-z}, \tag{30}$$

which can be solved using the Laplace transform (see Appendix E for details). For systems with $n_c = 2$, which have recently been shown to play an important role in the formation and proliferation of certain forms of amyloid fibrils,[16] the solution to Eq. (30) reads

$$\begin{aligned} C(z,\tau) = {} & e^{2z(1-e^{-\tau})} \left( \frac{1-z}{1-ze^{-\tau}} \right)^2 e^{-3\tau} C_{ss}(ze^{-\tau}) + \left( \frac{M(\infty)}{\sqrt{\xi}} - P(\infty) \right) e^{-3\tau} \left[ 1 - e^{2z(1-e^{-\tau})} \left( \frac{1-z}{1-ze^{-\tau}} \right)^2 \right] \\ & + \frac{P(\infty)}{2z^3} \left[ 2z^3 + 3z^2 - 3 - e^{2z(1-e^{-\tau})} \left( \frac{1-z}{1-ze^{-\tau}} \right)^2 (2z^3 e^{-3\tau} + 3z^2 e^{-2\tau} - 3) \right], \end{aligned} \tag{31}$$

where

$$C_{ss}(z) = \sum_{j=n_c}^{\infty} f_{ss}(j) z^j. \tag{32}$$

As a verification of the consistency of the generating function approach, we evaluate Eq. (31) at $z = 1$ and recover the expression for the polymer number concentration, Eq. (22),

$$C(z = 1, \tau) = P(\infty) + \left( \frac{M(\infty)}{\sqrt{\xi}} - P(\infty) \right) e^{-3\tau} = P(\tau). \tag{33}$$

Finally, expanding Eq. (31) in power series in $z$ yields the following closed-form expression for the length distribution

$$\begin{aligned} f(\tau, j) = {} & \left( \frac{M(\infty)}{\sqrt{\xi}} - P(\infty) \right) e^{-3\tau} A_j(\tau) + P(\infty) B_j(\tau) \\ & + e^{-3\tau} \left[ f_{ss}(j) e^{-j\tau} - \sum_{k=2}^{j-2} f_{ss}(k) e^{-k\tau} A_{j-k}(\tau) \right], \end{aligned} \tag{34}$$

where the coefficients $A_j(\tau)$ and $B_j(\tau)$ are given by

$$A_j(\tau) = \sum_{k=0}^{j-1} \frac{(1-e^{-\tau})^2 (k-j) + 1 - e^{-2\tau}}{k!} [2(1-e^{-\tau})]^k e^{-(j-k-2)\tau} - \frac{[2(1-e^{-\tau})]^j}{j!}, \tag{35}$$

$$\begin{aligned} B_j(\tau) = {} & \sum_{k=0}^{j-1} \frac{k - j + 4 + e^{-\tau}(j-k-2)}{2k!} [2(1-e^{-\tau})]^{k+1} e^{-(j-k+1)\tau} \\ & + \frac{3(j-1) + 3j(j-1)e^{-\tau} - 6j(j+1)e^{-2\tau} + 2j(j+1)e^{-3\tau}}{(j+3)(j+1)!} [2(1-e^{-\tau})]^j. \end{aligned} \tag{36}$$

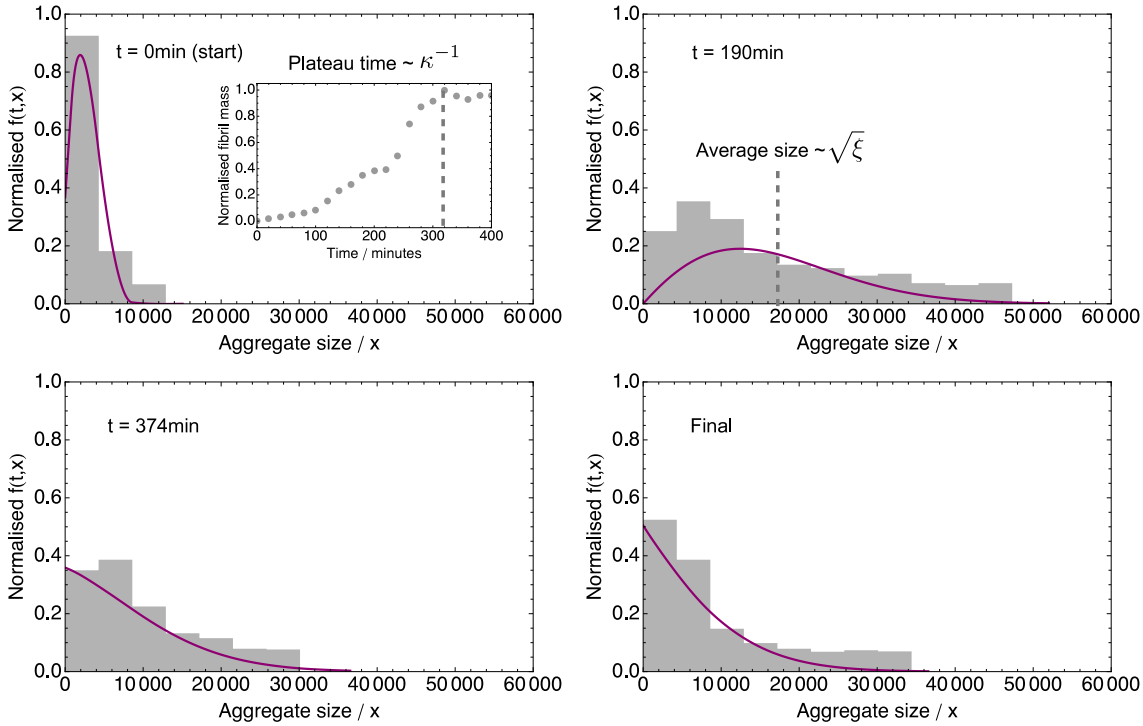


FIG. 4. Comparison between experimental measurements and analytical predictions for the length distribution at different times. The histograms represent the experimentally obtained length distributions of insulin filaments with histogram bin widths of 500 nm. The solid curves represent the predictions of Eqs. (11) and (34) with parameters: $k_+ = 2.9\times10^4$ M$^{-1}$ s$^{-1}$, $k_{\rm off} = 0$, $k_n = 0$, $k_- = 1.75\times10^{-8}$ s$^{-1}$, $n_c = 2$, $m(0) = 87$ $\mu$M, $M(0) = 87$ nM, $P(0) = 3.3\times10^{-11}$ M. We stress that the theoretical curves do not represent fits to the experimental data but are predictions using the values for the rate constants determined solely from the average length of aggregates at the stationary state ($\sim\sqrt{\xi}$) and the time $T$ at which $M(t)$ reaches the plateau.

### C. Discussion

Figure 3 shows the behaviour of the aggregate size distribution predicted by our analytical framework, Eqs. (11) and (34). As can be seen from the figure, the analytical integrated rate law obtained from the combination of Eqs. (11) and (34) accounts explicitly for the transition from an open-system behaviour at early times to a slow approach to steady state at later times. The aggregate distribution develops initially in the form of an advective front in aggregate size space until the stationary distribution Eq. (12) is reached. At later times, the filament distribution shifts towards shorter fragments as a result of fragmentation dominating over elongation. At steady-state $t = \infty$, contributions from exponential decaying terms in Eq. (34) vanish and $f(t,j)$ recovers the previously obtained steady-state distribution[27]

$$\frac{f(\infty,j)}{P(\infty)} = \frac{n_c[(n_c-1)n_c]^{j-n_c}(n_c-n_c^2+j+j^2)(n_c^2-1)!}{[1+(n_c-1)n_c+j]!}. \tag{37}$$

Importantly, the precise form of the length distribution reveals that these two stages of the reaction are controlled by different combinations of the rate parameters. During the first phase of the reaction, the time evolution and shape of the distribution are controlled by the parameters $\xi$, $\kappa$, and $\upsilon$: the parameter $\xi$ controls the shape of the distribution and describes a balance between filament elongation, which tends to make the aggregates longer, and filament breakage, which shifts the distribution towards smaller values of $j$; the temporal evolution of $f(t,j)$ is determined by the parameters $\kappa$ and $\upsilon$, whereby $\kappa$ gives the rate of aggregate multiplication and $\upsilon$ describes the wave-front propagation of $f(t,j)$ due to filament elongation. During the second, slower stage of the reaction, the time behaviour of the aggregate size distribution is controlled by the rate of fragmentation $k_-$. Note that at the end of the reaction, the fractional aggregate distribution, Eq. (37), is independent of any kinetic parameter.

## V. DYNAMICS OF FILAMENT LENGTH DISTRIBUTION REVEALS MOLECULAR MECHANISMS OF INSULIN ASSEMBLY

To date, the experimental analysis of filamentous growth reactions has mainly focussed on fitting average quantities, such as the polymer mass concentration $M(t)$.[43] Values for the governing rate-constants are estimated from a global fit of aggregate mass traces to analytically determined functions corresponding to the correct microscopic mechanism. While powerful, this approach is necessarily limited by the fact that it relies on the analysis of average quantities, which, in certain cases, cannot provide enough information for discriminating between different possible dominant microscopic mechanisms. For example, when the system proliferates through a monomer-independent secondary mechanism, fitting the time course of the aggregate mass only does not allow to establish whether the underlying microscopic mechanism is fibril breakage or saturated surface-catalyzed (secondary) nucleation.[44] We now show that the availability of our theoretical framework opens up new possibilities of studying filament length distributions quantitatively to complement the analysis of principal moments. We demonstrate the power of

this approach by analysing experimental measurements of the length distribution of growing insulin filaments. Fibril length distributions were measured from seeded aggregation kinetic experiments monitored by changes in Thioflavin T fluorescence and subsequent atomic force microscope (AFM) imaging at different times (see Appendix E for details on materials and methods). The experiment showed that starting from the initial filament distribution, filaments steadily grew longer through recruitment of free polypeptide molecules at the ends of aggregates until, after about 320 min, the length distribution shifted towards shorter fibrils. The experimental results for the length distribution of the insulin fibrils are plotted in Fig. 4. Importantly, while fitting of the time course for the polymer mass concentration $M(t)$ to kinetic models of protein filament formation dominated by fragmentation or saturated secondary nucleation would result in the same exact functional form,[44] the time evolution of the filament distribution is clearly consistent only with a fragmentation model, as saturated secondary nucleation would not lead to a shift of $f(t,j)$ towards shorter fibrils at late times. In order to compare our model predictions with the AFM data, the values for the rate constants for elongation and fragmentation are determined from the average length of aggregates at the stationary state ($\approx 1.7 \times 10^4$) and the time at which the mass concentration $M(t)$ of fibrils reaches the plateau ($\approx$320 min), yielding values of $k_+ = 2.9 \times 10^4$ $\mathrm{M}^{-1}$ $\mathrm{s}^{-1}$ and $k_- = 1.75 \times 10^{-8}$ $\mathrm{s}^{-1}$.[45] The prediction for the form of the filament length distribution using the obtained values for the rate constants is shown in Fig. 4. The obtained values for the kinetic parameters are in agreement with previous reports, considering changes in temperature.[46] Finally, we note that while a global fit of experimental measurements of the total aggregate mass concentration is able to fix only the combined rate parameter $k_+k_-$, the analysis of filament length distributions described here allows the determination of the individual rate constants. Overall, these results demonstrate the power of chemical kinetics for quantitatively predicting the behaviour of the fibril population with time hence providing a valuable tool to increase our mechanistic understanding of linear protein aggregation phenomena.

## VI. CONCLUSIONS

In this paper, we have presented an analytical study of the kinetics of breakable filament self-assembly. We have obtained closed-form expressions describing the temporal evolution of the length distribution of the resulting filamentous structures. These results uncover the basic physical behaviour that determines the time evolution of the aggregate size distribution. Furthermore, we have demonstrated that our theoretical model yields good agreement with experimental data of the length distribution of growing insulin fibrils and provides insights into the underlying microscopic mechanisms in action.

## ACKNOWLEDGMENTS

This work has been supported by St. John's College, Cambridge (T.C.T.M.), the Danish National Research Foundation (P.Y., M.H.J.), the Center for Models of Life (P.Y., M.H.J.), Magdalene College, Cambridge (A.K.B.), the Leverhulme Trust (A.K.B.), Elan Pharmaceuticals (A.K.B.), the Newman Foundation (T.P.J.K.), and BBSRC (T.P.J.K.).

## APPENDIX A: EXACT SOLUTION FOR THE LENGTH DISTRIBUTION IN AN OPEN SYSTEM

In this appendix, we focus on systems with constant monomer concentration. Building on the approach of Pöschel and Brilliantov,[22] we derive exact expressions describing the time evolution of the aggregate distribution in the discrete case, which is solution of the following master equation:

$$\frac{\partial f(\tau,j)}{\partial \tau} = \xi f(\tau,j-1) - \xi f(\tau,j) + 2\sum_{i=j+1}^{\infty} f(\tau,i) - (j-1)f(\tau,j) + \nu\delta_{j,n_c}, \tag{A1}$$

where we have introduced the rescaled time variable $\tau = k_- t$.

### 1. Principal moments

Preliminary insights into the form of the filament size distribution is obtained by considering the principal moments of the length distribution, defined as

$$I_n(t) = \sum_{j=n_c}^{\infty} j^n f(j,t). \tag{A2}$$

Of particular experimental interest are the aggregate number concentration $I_0(t) \equiv P(t)$ (zeroth moment) and the concentration of monomers in aggregates $I_1(t) \equiv M(t)$ (first moment). These are average quantities, which obey a set of two coupled ordinary differential equations (moment equations) obtained by summing master equation (A1) on both sides over aggregation number, yielding a solution as a sum of exponentials[16–22]

$$\begin{aligned} P(t) &= C_1 e^{\kappa t} + C_2 e^{-\kappa t} - \frac{n_c \nu}{\xi}, \\ M(t) &= \sqrt{\xi} C_1 e^{\kappa t} - \sqrt{\xi} C_2 e^{-\kappa t} - \nu, \end{aligned} \tag{A3}$$

where the constants $C_{1,2}$ are given by

$$C_{1,2} = \frac{1}{2}\left(P(0) \pm \frac{M(0)}{\sqrt{\xi}} \pm \frac{\nu}{\sqrt{\xi}}\right). \tag{A4}$$

### 2. Solution for the aggregate size distribution

In order to solve for $f(\tau,j)$, we rewrite the first fragmentation term in Eq. (A1) as

$$\sum_{i=j+1}^{\infty} f(\tau,i) = P(\tau) - f(\tau,j) - f(\tau,j-1) - \sum_{i=n_c}^{j-2} f(\tau,i), \tag{A5}$$

such that from Eq. (A3), Eq. (A1) becomes

$$\begin{aligned} \frac{\partial f(\tau,j)}{\partial \tau} = &-(\xi + j + 1)f(\tau,j) + (\xi - 2)f(\tau,j-1) \\ &- 2\sum_{i=n_c}^{j-2} f(\tau,i) + 2C_1 e^{\sqrt{\xi}\tau} \\ &+ 2C_2 e^{-\sqrt{\xi}\tau} - 2P + \nu\delta_{j,n_c}, \end{aligned} \tag{A6}$$

where $P = n_c\nu/\xi$. Written in this form, Eq. (A6) can be solved recursively, as the time evolution of $f(\tau, j)$ is now coupled only to that of the concentrations $f(\tau, i)$ with $i < j$. For $j = n_c$, Eq. (A6) admits the simple solution[22]

$$f(\tau, n_c) = \frac{2C_1}{\xi + \sqrt{\xi} + n_c + 1} e^{\sqrt{\xi}\tau} + \frac{2C_2}{\xi - \sqrt{\xi} + n_c + 1} e^{-\sqrt{\xi}\tau} + C_{n_c} e^{-\alpha_{n_c}\tau} - \frac{2P - \nu}{\xi + n_c + 1}, \tag{A7}$$

where $\alpha_k = \xi + k + 1$ and the constant of integration

$$C_{n_c} = f_0(n_c) - \frac{2C_1}{\xi + \sqrt{\xi} + n_c + 1} - \frac{2C_2}{\xi - \sqrt{\xi} + n_c + 1} + \frac{2P - \nu}{\xi + n_c + 1} \tag{A8}$$

has been determined by implementing the initial condition at $t = 0$. Since we are interested in the long-time limit, we seek for a general solution of Eq. (A6) in the form[21]

$$f(\tau, j) = A_j e^{\sqrt{\xi}\tau}, \tag{A9}$$

where the coefficients $A_j$ obey the following recursion relation, which are obtained by inserting Eq. (A9) into Eq. (A6):[22]

$$\frac{A_j}{C_1} = \frac{(\xi - 2)}{\xi + \sqrt{\xi} + j + 1} \frac{A_{j-1}}{C_1} + \frac{2}{\xi + \sqrt{\xi} + j + 1} \left(1 - \sum_{i=n_c}^{j-2} \frac{A_i}{C_1}\right). \tag{A10}$$

### 3. Solution of the recursion relation

As an intermediate step towards the solution of Eq. (A1), now derive the exact solution of the following recursion relation:

$$x_j = \frac{a-2}{b+j+1} x_{j-1} + \frac{2}{b+j+1} \left(1 - \sum_{i=n_c}^{j-2} x_i\right), \tag{A11}$$

subject to given initial conditions $x_{n_c} =: X$ and $x_{n_c+1} = [(a-2)X + 2]/(b + n_c + 2) =: Y$. An obstacle to solving Eq. (A11) is the unlimited number of terms involved in the sum. To avoid this difficulty, we transform the recurrence relation into a recursion of fixed degree by setting up a subtraction of sums for $j \geq n_c + 2$,

$$(b + j + 1)x_j = (a - 2)x_{j-1} + 2\left(1 - \sum_{i=n_c}^{j-2} x_i\right), \tag{A12}$$

$$(b + j)x_{j-1} = a x_{j-2} + 2\left(1 - \sum_{i=n_c}^{j-2} x_i\right). \tag{A13}$$

Productively, subtracting Eq. (A12) from Eq. (A13) yields

$$(b + j + 1)x_j = (a + b - 2 + j)x_{j-1} - a x_{j-2}. \tag{A14}$$

By using the ansatz $x_j = \tilde{x}_j/(b + n_c + 1)_{j+1-n_c}$, where $(a)_n = a(a+1)\ldots(a+n-1)$ denotes the Pochhammer symbol, and by writing $j = k + n_c$ with $k = 0, 1, \ldots$, we can recast Eq. (A14) into the simpler form

$$\tilde{x}_k = (a + b - 2 + n_c + k)\tilde{x}_{k-1} - a(b + n_c + k)\tilde{x}_{k-2} \tag{A15}$$

valid for all $k \geq 2$. To solve Eq. (A15), we introduce the generating function

$$C(z) = \sum_{k=2}^{\infty} z^k \tilde{x}_k. \tag{A16}$$

Multiplying Eq. (A15) with $z^k$ and taking the sum over $k$ on both sides yields a differential equation for $C(z)$,

$$(az^3 - z^2)\frac{\partial C(z)}{\partial z} = \left[(a + b - 1 + n_c)z - a(b + n_c + 2)z^2 - 1\right] C(z) + \left[(a + b + n_c)\tilde{x}_1 - a(b + n_c + 2)\tilde{x}_0\right] z^2 - a(b + n_c + 3)\tilde{x}_1 z^3, \tag{A17}$$

where $\tilde{x}_0 = (b + n_c + 1)X$ and $\tilde{x}_1 = (b + n_c + 1)(b + n_c + 2)Y$. According to Eq. (A16), to obtain $\tilde{x}_k$, we solve Eq. (A17) subject to the initial condition $C(z = 0) = 0$,

$$C(z) = \frac{az^2(A + Bz + Cz^2) + Dze^{-1/z}E_{b+n_c+1}(-1/z)}{(b + n_c + 1)(az - 1)^3}, \tag{A18}$$

where

$$A = a(2b - a + 2n_c + 4)\tilde{x}_0 + (a - 3b - 3n_c - 3)\tilde{x}_1, \tag{A19}$$

$$B = a(1 + b + n_c)(3\tilde{x}_1 - a\tilde{x}_0), \tag{A20}$$

$$C = -a^2(1 + b + n_c)\tilde{x}_1, \tag{A21}$$

$$D = a[2(2 + b + n_c) - (1 + b + n_c)(2 + b + n_c) - a^2]\tilde{x}_0 + [a^2 + (b + n_c - 2a)(1 + b + n_c)]\tilde{x}_1 \tag{A22}$$

and subsequently expand the resulting generating function $C(z)$ in power series in $z$ around the point $z = 0$, yielding

$$x_j = \frac{\Gamma(b+n_c+1)}{(1+b+n_c)\Gamma(b+j+2)}\Big[D\sum_{s=0}^{n}\frac{a_m}{2}(m+1)(m+2)\frac{\Gamma(1+b+n_c+n-m)}{\Gamma(b+n_c+1)} - A\frac{a^{n+1}}{2}(n+1)(n+2) - B\frac{a^n}{2}n(n+1) - C\frac{a^{n-1}}{2}n(n-1)\Big], \quad \text{(A23)}$$

where $n = j - n_c - 2$.

### 4. Discussion

Using Eq. (A23), we are now in the position to obtain explicitly the following exact solution for the aggregate distribution:

$$\frac{f(t,j)}{P(t)} = a_m\frac{\xi^{m-1}}{c}\frac{\Gamma(c)}{\Gamma(c+m+1)} + \frac{\alpha}{c}\sum_{s=0}^{m-2}\xi^s(s+1)(s+2) \times \frac{\Gamma(c+m-s-2)}{\Gamma(c+m+1)}, \quad \text{(A24)}$$

with $m = j - n_c$, $c = \xi + \sqrt{\xi} + n_c + 1$, $d = \xi - \sqrt{\xi} + n_c + 1$, and

$$a_m = 3m(m+1)(n_c-1) + m(m+1)(n_c+2)\sqrt{\xi} \quad \text{(A25)}$$
$$+ 2\xi[3 + m(m+n_c+2)] + (m+1)(m+2)\xi^{3/2}$$
$$+ 2(m+1)\xi^2, \quad \text{(A26)}$$
$$b = 6(n_c-1) + (1+5n_c)\sqrt{\xi} + 5\xi + 3\xi^{3/2}. \quad \text{(A27)}$$

Computation of the principal moments using Eq. (A24) yields

$$P(t) = \sum_{j=n_c}^{\infty} f(t,j) = C_1 e^{\kappa t} \quad \text{(A28)}$$

and

$$M(t) = \sum_{j=n_c}^{\infty} j f(t,j) = [(2n_c-1) + \sqrt{\xi}]C_1 e^{\kappa t}. \quad \text{(A29)}$$

It is interesting to note the appearance in Eq. (A29) of additional terms proportional to $(2n_c - 1)C_1$ when compared to Eq. (A3). This difference originates from the fact that the derivation of Eq. (A3) neglects the term $-(2n_c - 1)k_- P(t)$ describing the production of monomers when a filament breaks at a position that is closer than $(n_c - 1)$ bonds from either fibril ends. This approximation is justified for $\xi \gg 1$, in which case Eq. (A29) becomes

$$M(t) \approx \sqrt{\xi} C_1 e^{\kappa t} \quad \text{(A30)}$$

and recovers Eq. (A3) exactly.

## APPENDIX B: CONTINUUM-LIMIT DESCRIPTION OF BREAKABLE FILAMENT ASSEMBLY

In this appendix, we provide the details pertaining to the derivation of continuum-limit master equation (2).

In the transformation from a discrete to a continuum description of protein aggregation, we replace the discrete index $j$ with a continuum variable $x$ and expand the finite differences in Eq. (1) in terms of partial derivatives to leading order

$$f(t, j \pm 1) = f(t,x) \pm \frac{\partial f(t,x)}{\partial x} + O(f''), \quad \text{(B1)}$$

and replace sums with integrals, e.g.,

$$\sum_{i=j+1}^{\infty} f(t,i) \approx \int_x^{\infty} f(t,z)dz. \quad \text{(B2)}$$

By introducing the polymerisation drift coefficient $v(t) = 2[k_+ m(t) - k_{\text{off}}]$, the master equation, Eq. (1), can be therefore formulated in the continuum limit as[27,34]

$$\frac{\partial f(t,x)}{\partial t} = -v(t)\frac{\partial f(t,x)}{\partial x} - k_- x f(t,x) + 2k_- \int_x^{\infty} f(t,z)dz + k_n m(t)^{n_c}\delta(x - n_c). \quad \text{(B3)}$$

We note that in the transition from a discrete to a continuum formulation of Eq. (1), the multiplicative pre-factor $(x - 1)$ in the loss term related to fragmentation has been replaced by $x$. This fact follows intuition since rewriting $(x - 1) \approx x$ is justified for large aggregation numbers and also because this approximation corresponds to a scenario in which filaments can break anywhere along the continuous chain, even infinitely close to the ends. We also note that the continuum approximation provided by Eq. (B3) was derived by replacing finite differences with first-order derivatives. In general, a better approximation is obtained by considering higher order terms in the Taylor expansion

$$f(t, j \pm 1) = f(t,x) \pm \frac{\partial f(t,x)}{\partial x} + \frac{1}{2}\frac{\partial^2 f(t,x)}{\partial x^2} + \cdots. \quad \text{(B4)}$$

For example, if derivatives up to second order are considered in the expansion, an additional term on the right hand side of the continuum master equation is obtained, which is of the form

$$D(t)\frac{\partial^2 f(t,x)}{\partial x^2}, \quad \text{(B5)}$$

where $D(t) = k_+ m(t) + k_{\text{off}}$ describes a time dependent diffusion process in length space that acts to broaden the length distribution.

## APPENDIX C: FILAMENT SIZE DISTRIBUTION IN THE PRESENCE OF AGGREGATE DEGRADATION

In this appendix, we extend the theoretical framework of Sec. III to the situation when aggregate degradation mechanisms are active.[21] Aggregates are assumed to undergo degradation at the rate $k_{\text{deg}}$, which is independent of their size; this framework, however, could in principle accommodate a size dependent rate of degradation. Under these circumstances,

continuum-limit master equation (2) becomes

$$\frac{\partial f(\tau,x)}{\partial \tau} = -\xi\frac{\partial f(\tau,x)}{\partial x} - (x+\alpha)f(\tau,x) + 2\int_x^\infty f(\tau,z)dz + \nu\delta(x-n_c), \tag{C1}$$

where $\tau = k_- t$ and

$$\xi = \frac{2[k_+ m(0) - k_{\text{off}}]}{k_-}, \quad \nu = \frac{k_n m(0)^{n_c}}{k_-}, \quad \alpha = \frac{k_{\text{deg}}}{k_-}. \tag{C2}$$

Using the same arguments as in Sec. III, we introduce the Laplace transform

$$\hat{f}(s,x) = \int_0^\infty f(\tau,x)e^{-s\tau}d\tau \tag{C3}$$

and obtain after differentiation with respect to $x$,

$$\xi\frac{\partial^2 \hat{f}(s,x)}{\partial x^2} + (x+s+\alpha)\frac{\partial \hat{f}(s,x)}{\partial x} + 3\hat{f}(s,x) = 0 \tag{C4}$$

subject to the boundary condition

$$\hat{f}(s, x=n_c) = \frac{\nu}{\xi s}. \tag{C5}$$

The solution is

$$\hat{f}(s,x) = \frac{\nu e^{-\frac{(x-n_c)(x+n_c+2s+2\alpha)}{2\xi}}[(x+s+\alpha)^2 - \xi]}{\xi s[(n_c+s+\alpha)^2 - \xi]}. \tag{C6}$$

The inverse Laplace transform of Eq. (C6) yields the following solution for the aggregate size distribution:

$$f(t,x) = \frac{\nu e^{-\frac{x^2-n_c^2}{2\xi}}}{\xi(\xi-\alpha^2)}\left\{[\xi-(x+\alpha)^2]\,e^{-\frac{\alpha(x-n_c)}{\xi}} + x(x+2\alpha)\cosh\left(\kappa t + \frac{n_c-x}{\sqrt{\xi}}\right)e^{-k_{\text{deg}}t} + x\left(2\sqrt{\xi} + \frac{\alpha x}{\sqrt{\xi}}\right)\sinh\left(\kappa t + \frac{n_c-x}{\sqrt{\xi}}\right)e^{-k_{\text{deg}}t}\right\} H(vt - x + n_c), \tag{C7}$$

where we used $\sqrt{\xi} \gg n_c$. The long-time limit of Eq. (C7) reads

$$\frac{f(t,x)}{P(t)} \overset{\kappa t \gg 1}{\to} \frac{e^{-\frac{(x-n_c)(2\sqrt{\xi}+x+n_c)}{2\xi}}x(x+2\sqrt{\xi})}{\xi^{3/2}}, \tag{C8}$$

where

$$P(t) = \frac{\nu}{2(\sqrt{\xi}-\alpha)}e^{(\kappa-k_{\text{deg}})t}, \quad \kappa t \gg 1. \tag{C9}$$

It is worth noting that in the presence of degradation, the form of the stationary fractional occupation, Eq. (C8), is the same as that of Eq. (12).

## APPENDIX D: PRESENCE OF SEED MATERIAL

The approach outlined in Sec. III can be generalized to the case when a non-zero concentration of seed aggregates is present initially. In this case, the general solution for $f(t,x)$ can be expressed as the sum of the solution of the corresponding homogeneous equation, Eq. (11), and a particular solution of the non-homogeneous equation. For example, for the initial condition

$$f(t=0,x) = P(0)\delta(x-n_c), \tag{D1}$$

the Laplace transform of the non-homogeneous solution, $\hat{f}(s,x) = \int_0^\infty f(\tau,x)e^{-s\tau}d\tau$ with $\tau = k_- t$, satisfies

$$s\hat{f}(s,x) - P(0)\delta(x-n_c) = -\xi\frac{\partial \hat{f}(s,x)}{\partial x} - x\hat{f}(s,x) + 2\int_x^\infty \hat{f}(s,z)dz. \tag{D2}$$

Following a similar reasoning as in Sec. III, the required solution to Eq. (D2) reads

$$\hat{f}(s,x) = \frac{P(0)e^{-\frac{(x-n_c)(x+n_c+2s)}{2\xi}}[(x+s)^2-\xi]}{\xi[(n_c+s)^2-\xi]}. \tag{D3}$$

Computing the inverse Laplace transform of Eq. (D3) for $\sqrt{\xi} \gg n_c$ yields

$$f(t,x) = \delta(x-vt-n_c)e^{-\frac{x^2-(x-vt)^2}{2\xi}} + f_h(t,x) + \frac{P(0)e^{-\frac{x^2-n_c^2}{2\xi}}x\left[2\sqrt{\xi}\cosh\left(\kappa t + \frac{n_c-x}{\sqrt{\xi}}\right) + x\sinh\left(\kappa t + \frac{n_c-x}{\sqrt{\xi}}\right)\right]}{\xi^{3/2}}H(vt-x+n_c), \tag{D4}$$

where $f_h(t,x)$ is the expression Eq. (11). Equation (D4) for the time evolution of the aggregate size distribution has three terms. The first term is the initial distribution that displaces through size space at velocity $v$ and decays exponentially as it moves along the $x$-axis. The advective motion in size space is due to filament elongation, while the decay is due to the combined effects of elongation and fragmentation. The second and third terms in Eq. (D4) represent the exponentially growing

biased Gaussian concentration profile of the aggregate distribution that trails behind the travelling initial delta function. As $t \to \infty$, the moving initial distribution vanishes and Eq. (D4) recovers a stationary behaviour similar to Eq. (12) independently of the initial state of the assembly reaction

$$\frac{f(t,x)}{P(t)} \overset{\kappa t \gg 1}{\to} \frac{e^{-\frac{(x-n_c)(x+n_c+2\sqrt{\xi})}{2\xi}} x(x+2\sqrt{\xi})}{\xi^{3/2}}, \tag{D5}$$

where from Eq. (D4),

$$P(t) = \frac{\nu}{2\sqrt{\xi}}\left(e^{\kappa t} - e^{-\kappa t}\right) + \frac{P(0)}{2}\left(e^{\kappa t} + e^{-\kappa t}\right) - \frac{n_c \nu}{\xi}. \tag{D6}$$

## APPENDIX E: DERIVATION OF EQ. (34)

In this appendix, we provide the mathematical details pertaining to the solution of Eq. (30). Introducing the Laplace transform

$$\hat{C}(z,s) = \int_0^\infty C(z,\tau) e^{-s\tau} d\tau, \tag{E1}$$

transforms Eq. (30) into

$$z\frac{\partial \hat{C}(z,s)}{\partial z} = \left[n_c(n_c-1)(z-1) - s + 1 - \frac{2}{1-z}\right]\hat{C}(z,s) + \frac{2z^{n_c}\hat{P}(s)}{1-z} + C_{ss}(z), \tag{E2}$$

where

$$\hat{P}(s) = \frac{P(\infty)}{s} + \frac{\frac{M(\infty)}{\sqrt{\xi}} - P(\infty)}{s + 2n_c - 1}. \tag{E3}$$

The solution of Eq. (E2) subject to $\hat{C}(z=0,s) = 0$ reads

$$\hat{C}(z,s) = \frac{e^{az}(1-z)^2}{z^{a+1}} \int_0^z \frac{w^a e^{-aw}}{(1-w)^2}\left(\frac{w}{z}\right)^s \times \left[C_{ss}(w) + \frac{2w^{n_c}\hat{P}(s)}{1-w}\right] dw, \tag{E4}$$

where, for convenience, we have introduced

$$a = n_c(n_c - 1). \tag{E5}$$

Taking the inverse Laplace transform of Eq. (E4) yields

$$C(z,\tau) = e^{az(1-e^{-\tau})}\left(\frac{1-z}{1-ze^{-\tau}}\right)^2 e^{-(a+1)\tau} C_{ss}(ze^{-\tau}) + \frac{2P(\infty)e^{az}(1-z)^2}{z^{a+1}} \int_{ze^{-\tau}}^z \frac{w^{n_c^2}}{(1-w)^3} e^{-aw} dw + 2\left(\frac{M(\infty)}{\sqrt{\xi}} - P(\infty)\right) e^{-(2n_c-1)\tau} \frac{e^{az}(1-z)^2}{z^{a+2-2n_c}} \int_{ze^{-\tau}}^z \frac{w^{(n_c-1)^2}}{(1-w)^3} e^{-aw} dw, \tag{E6}$$

where we used the following results:

$$\mathcal{L}^{-1}\left[\left(\frac{w}{z}\right)^s\right] = \delta(\tau + \log w - \log z) = \delta\left(w - ze^{-\tau}\right) w, \tag{E7}$$

$$\mathcal{L}^{-1}\left[s^{-1}\left(\frac{w}{z}\right)^s\right] = H\left(w - ze^{-\tau}\right), \tag{E8}$$

$$\mathcal{L}^{-1}\left[(s+q)^{-1}\left(\frac{w}{z}\right)^s\right] = e^{-q\tau}\left(\frac{z}{w}\right)^q H\left(w - ze^{-\tau}\right). \tag{E9}$$

Evaluating integrals in Eq. (E6) for $n_c = 2$ yields Eq. (34) of the main text. In addition, at steady state, Eq. (E6) becomes

$$\frac{C(z,\infty)}{P(\infty)} = \frac{2e^{az}(1-z)^2}{z^{a+1}} \int_0^z \frac{w^{a+n_c} e^{-aw}}{(1-w)^3} dw. \tag{E10}$$

Expanding Eq. (E10) as a power series in $z$ yields the aggregate size distribution at steady state[27]

$$\frac{f(\infty,j)}{P(\infty)} = \frac{n_c[(n_c-1)n_c]^{j-n_c}(n_c - n_c^2 + j + j^2)(n_c^2 - 1)!}{[1 + (n_c-1)n_c + j]!}. \tag{E11}$$

## APPENDIX F: MATERIALS AND METHODS

### 1. Fluorescence measurements

Fibril length distributions were measured from seeded fibril growth kinetic runs monitored by changes in Thioflavin T fluorescence. Bovine insulin monomer (Gemini Bio-Products, USA) was dissolved at a concentration of 0.5 mg ml$^{-1}$ in a solution of 10 mM HCl and 30 mM NaCl in Milli-Q water with 60 $\mu$M Thioavin T (ThT) fluorescent dye and 0.5 $\mu$g ml$^{-1}$ seed fibrils at a total volume of 1.5 ml. Precise protein monomer concentration was determined by absorbance spectroscopy using a Cary 400 Scan UV-visible spectrophotometer or Thermo Scientific Nanodrop 2000 spectrophotometer with molar extinction coefficient 1.0 cm$^{-1}$ for 1.0 mg ml$^{-1}$ at 276 nm. Immediately before starting the kinetic run, seed fibres were added to the cuvettes containing the monomer solution and mixed by inversion several times. Thermostat-controlled temperature was maintained at 45 °C and the progress of fibril growth was followed by measuring ThT fluorescence at 1 min intervals with excitation wavelength 440 nm and emission wavelength 480 nm. After each 30 min period, the reaction mixture was mixed by inverting the cuvette several times before a 10 $\mu$l reaction aliquot was removed and diluted in 10 mM HCl to appropriate concentrations for AFM imaging. Both the reaction aliquots and reaction aliquot dilutions were stored at 4 °C until preparation of AFM slides to slow fibril growth and breakage kinetics. A fluorescence measurement and reaction aliquot were also taken after 24 h to provide information about the long-time limit at equilibrium. Taken as a value corresponding to 100% completion of aggregation, this final fluorescence value was used to normalise the

fluorescence measurements of each sample to give the fraction of protein incorporated into filaments as a function of time. This allowed sample comparison and the extent of the aggregation process at each time point, when a reaction aliquot was removed, to be determined. As suggested by DePace *et al.*,[47] to avoid shear forces causing fibril fragmentation during pipetting, wide-bore pipette tips were used by cutting all pipette tips 1 cm from the end to create a width 1-2 mm in diameter. Furthermore, mixing of solutions during dilution was achieved only by inversion to avoid shearing during all manipulations.

### 2. AFM imaging

20 $\mu$l of protein sample was deposited onto a cleaved mica surface affixed onto a glass microscope slide and left to air dry before subsequently washing with 20 $\mu$l Milli-Q water and blotting dry to remove any deposited salt which would obscure AFM imaging. Samples were imaged using intermittent contact tapping mode AFM in air using a JPK Nanowizard II AFM (JPK Instruments AG, Germany) with cantilever tips supplied by MikroMasch, Estonia (resonance frequencies 75, 105, and 155 kHz and force constants 0.6, 0.95, and 1.75 N m$^{-1}$, respectively). After initial AFM imaging, insulin fibril samples with a high background monomer concentration were covered with 30 $\mu$l 10 mM HCl and left for 5 min to redissolve deposited monomer before blotting. Samples were then washed twice with 20 $\mu$l Milli-Q water and blotted dry before reimaging. Height trace data from AFM images was processed using either JPK Data Processing or Gwyddion image processing software for vertical rescaling. Data levelling for background subtraction and correction for an offset in height data was achieved by applying a global linear plane subtraction from all data to remove any tilt angle from the sample support, or, where appropriate, correction by subtraction of a quadratic or quartic polynomial fit from each scan line independently.

### 3. Fibril length measurements

AFM images were analysed using ImageJ software to measure fibril contour lengths. Fibril curvature was approximated by using a segmented straight line tool. Fibrils of indeterminate length extending outside the image region and cutoff by the image border were not included.


[1]F. Oosawa and S. Asakura, *Thermodynamics of the Polymerisation of Proteins* (Academic Press, Waltham, MA, USA, 1975).
[2]F. Oosawa and M. Kasai, J. Mol. Biol. **4**, 10 (1962).
[3]F. Oosawa, J.Theor. Biol. **27**, 69–86 (1970).
[4]L. S. Tobacman and E. D. Korn, J. Biol. Chem. **258**, 3207 (1983).
[5]C. Frieden and D. W. Goddette, Biochemistry **22**, 5836 (1983).
[6]A. Wegner and J. Engel, Biophys. Chem. **3**, 215 (1975).
[7]C. M. Dobson, Nature **426**, 884 (2003).
[8]F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).
[9]S. R. Collins, A. Douglas, R. D. Vale, and J. S. Weissman, PLoS Biol. **2**, e321 (2004).
[10]M. Tanaka, S. R. Collins, B. H. Toyama, and J. S. Weissman, Nature **442**, 585 (2006).
[11]W. F. Xue, S. W. Homans, and S. E. Radford, Proc. Natl. Acad. Sci. U. S. A. **105**, 8926 (2008).
[12]W.-F. Xue, A. L. Hellewell, W. S. Gosal, S. W. Homans, E. W. Hewitt, and S. E. Radford, J. Biol. Chem. **284**, 34272 (2009).
[13]Y.-Q. Wang, A. K. Buell, X.-Y. Wang, M. E. Welland, C. M. Dobson, T. P. J. Knowles, and S. Perrett, J. Biol. Chem. **286**, 12101 (2011).
[14]W. F. Xue and S. E. Radford, Biophys. J. **105**, 2811 (2013).
[15]W. Lee, H. Jung, M. Son, H. Lee, T. J. Kwak, G. Lee, C. H. Kim, S. W. Lee, and D. S. Yoon, RSC Adv. **4**, 56561 (2014).
[16]T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, Science **326**, 1533 (2009).
[17]S. I. A. Cohen, M. Vendruscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, and T. P. J. Knowles, J. Chem. Phys. **135**, 065105 (2011).
[18]F. A. Ferrone, J. Hofrichter, and W. A. Eaton, J. Mol. Biol. **183**, 611 (1985).
[19]F. Ferrone, Methods Enzymol. **309**, 256 (1999).
[20]M. F. Bishop and F. A. Ferrone, Biophys. J. **46**, 631 (1984).
[21]J. Masel, V. A. Jansen, and M. A. Nowak, Biophys. Chem. **77**, 139 (1999).
[22]T. Pöschel, N. V. Brilliantov, and C. Froemmel, Biophys. J. **85**, 3460 (2003).
[23]T. C. T. Michaels and T. P. J. Knowles, Int. J. Mod. Phys. B **29**, 1530002 (2015).
[24]D. M. Walsh, I. Klyubin, J. V. Fadeeva, W. K. Cullen, R. Anwyl, M. S. Wolfe, M. J. Rowan, and D. J. Selkoe, Nature **416**, 535 (2002).
[25]M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, and M. Stefani, Nature **416**, 507 (2002).
[26]R. Kayed, E. Head, J. L. Thompson, T. M. McIntire, S. C. Milton, C. W. Cotman, and C. G. Glabe, Science **300**, 486 (2003).
[27]S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, J. Chem. Phys. **135**, 065107 (2011).
[28]P. Arosio, M. Beeg, L. Nicoud, and M. Morbidelli, Chem. Eng. Sci. **78**, 2132 (2012).
[29]T. Medkour, F. Ferrone, F. Galacttros, and P. Hannaert, Acta Biotheor. **56**, 103 (2008).
[30]T. C. T. Michaels and T. P. J. Knowles, Am. J. Phys. **82**, 476 (2014).
[31]A. Cacciuto, S. Auer, and D. Frenkel, Nature (London) **428**, 404 (2004).
[32]G. A. Garcia, S. I. A. Cohen, C. M. Dobson, and T. P. J. Knowles, Phys. Rev. E. **89**, 032712 (2014).
[33]T. L. Hill, Biophys. J. **44**, 285 (1983).
[34]H. Flyvbjerg, T. E. Holy, and S. Leibler, Phys. Rev. E **54**, 5538 (1996).

[35]To see why the integral of the second term in Eq. (6) is divergent, it is convenient to exploit the following asymptotic expansion: $D(x) \sim \frac{1}{2x}$ for $x \to \infty$. This formula makes it easy to see that $\int_{n_c}^{\infty} D(\frac{x+s}{\sqrt{\xi}})dx \sim \int_{n_c}^{\infty} \frac{\sqrt{\xi}}{2(x+s)}dx$ is divergent.
[36]Note that the condition $\sqrt{\xi} \gg n_c$ is necessary to ensure that long filaments are present in the system.
[37]Numerical realizations of Eq. (1) were generated using a 4th order Runge–Kutta differential equation solving routine. The infinite number of equations in (1) requires truncation of the master equation at some finite maximum size $N$ of filaments. For the set of parameters used in Fig. 2, $N \sim 10^6$ was chosen to be much bigger than the average size of aggregates at steady-state ($\sim\sqrt{\xi}$), such that truncation errors are negligible. Alternative methods include numerical simulation of Eq. (1) using dynamic Monte Carlo.[22]
[38]Because under typical conditions $\xi \gg 1$, we note that Eq. (12) is equivalent to the previously derived stationary solution (Eq. (36) in Ref. 27).

[39]T. C. T. Michaels, G. A. Garcia, and T. P. J. Knowles, J. Chem. Phys. **140**, 194906 (2014).
[40]P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, UK, 2010).

[41]The generating function $C(z, \tau)$ represents a particularly convenient way of encoding the system dynamics into a single expression, whose coefficients correspond to the aggregate size distribution $f(\tau, j)$. Generating functions are intimately connected with the moments of the filament distribution. For example, the aggregate number and mass concentrations can be obtained from the knowledge of $C(z, \tau)$ as follows:

$$C(z = 1, \tau) = P(\tau), \quad \left.\frac{\partial C(z, \tau)}{\partial z}\right|_{z=1} = M(\tau).$$

[42]In the presence of seed material at the beginning of the reaction, one can use the same theoretical framework but the expression Eq. (20) becomes $T = 1/\kappa \log[M(\infty)/(\sqrt{\xi}C_1)]$, where $C_1 = 1/2(P(0) + M(0)/\sqrt{\xi} + \nu/\sqrt{\xi})$.

[43]S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, J. Mol. Biol. **421**, 160 (2012).
[44]G. Meisl, X. Yang, E. Hellstrand, B. Frohm, J. B. Kirkegaard, S. I. A. Cohen, C. M. Dobson, S. Linse, and T. P. J. Knowles, Proc. Natl. Acad. Sci. U. S. A. **111**, 9384 (2014).
[45]From a measurement of the average length of aggregates at stationary state, we can determine $\sqrt{\xi} \approx 17\,000$. Using this result, we can fix the rate parameter $\kappa$ from a measurement of the time at which $m(t)$ reaches the plateau: $T = 1/\kappa \log(2m_{\mathrm{tot}}/P(0)\sqrt{\xi})$. This yields $\kappa \approx 3.0 \times 10^{-4}$ s$^{-1}$. Combining the results for $\sqrt{\xi}$ and $\kappa$, we arrive at the values for $k_+$ and $k_-$ reported in the main text.
[46]J. F. Smith, T. P. J. Knowles, C. M. Dobson, C. E. Macphee, and M. E. Welland, Proc. Natl. Acad. Sci. U. S. A. **103**, 15806 (2006).
[47]A. H. DePace and J. S. Weissman, Nat. Struct. Biol. **9**, 389 (2002).