Nucleation-conversion-polymerization reactions of biological macromolecules with prenucleation clusters

Gonzalo A. Garcia, Samuel I. A. Cohen, Christopher M. Dobson, and Tuomas P. J. Knowles*
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
(Received 29 January 2014; published 24 March 2014)

The self-assembly of biomolecules, such as peptides and proteins, into filaments is conventionally understood as a nucleated polymerization reaction. However, detailed analysis of experimental observation has revealed recently that nucleation pathways generate growth-competent nuclei via a cascade of metastable intermediate species, which are omitted in conventional models of filamentous growth based on classical nucleation theory. Here we take an analytical approach to generalizing the classical theory of nucleated polymerization to include the formation of these prenucleation clusters, providing a quantitative general classification of the behavior exhibited by these nucleation-conversion-polymerization reactions. A phase diagram is constructed, and analytical predictions are derived for key experimental observables. Using this approach, we delineate the characteristic time scales that determine the nature of biopolymer growth phenomena.

DOI: 10.1103/PhysRevE.89.032712 PACS number(s): 87.15.hg, 82.39.−k, 87.10.−e, 87.15.R−

I. INTRODUCTION

The self-assembly processes by which macromolecules combine to form filaments have been explored extensively over the past 5 decades [1–10], both in the context of normal biology [1,2,11] and in relation to a diverse class of human disorders, including protein misfolding diseases [12–20], prion conditions [21–25], and sickle-cell anemia [4,26,27]. As with other complex dynamical systems, a critical advance in understanding the microscopic details of these self-assembly reactions has been the development of kinetic rate laws in the form of master equations [2,7,28]. These microscopic rate laws have provided an increasingly detailed understanding of experimental self-assembly data [2,28,29] in terms of processes defined on the molecular level. These processes include primary [2] and secondary [4,30] nucleation pathways which create new aggregates and the fibril elongation process [31] which results in their growth. With recent advances in single-molecule detection techniques, several recent experimental reports [32–34] have revealed that the primary nucleation pathways in many important biological self-assembly reactions are characterized by nonclassical kinetics. In particular, nucleation processes are observed to populate metastable intermediates that undergo further structural conversions before possessing the ability to grow into elongated filaments. The formation of these prenucleation clusters is not accounted for by the kinetic rate laws of nucleated polymerization. Although particular cases have been studied [32,35,36], a global classification of the behavior exhibited by these systems has not been developed.

In the present work, we generalize the theory of classical nucleated polymerization to include these *cascade nucleation* processes and address two critical questions that determine the nature of the self-assembly process. First, under what conditions do these aggregating systems, defined in Sec. II, form elongated filaments (Sec. III)? Second, in which cases and in what ways do the additional conversion reactions influence macroscopic observables (Secs. IV–VII)? Using analytical techniques, we address both of these questions quantitatively, providing a phase diagram (Sec. VIII, Appendix) in terms of the key physical time scales present in such systems. In addition to the classically understood regions of parameter space, several new and distinct growth regimes emerge. Finally, Sec. IX considers the nucleation-conversion-polymerization pathway as an alternative to secondary pathways for describing high-order emergence of fibril mass with time.

II. THE GENERAL CASCADE NUCLEATION SYSTEM

The cascade nucleation process is summarized schematically in Fig. 1; the conversion steps, which are not accounted for in classical descriptions of nucleated polymerization, are illustrated in the yellow horizontal shaded band. These steps represent conformational changes, for example, associated with alternations in the internal degrees of freedom of the protein molecules [32]. The rate constant for the formation of the first cluster is k_n with a reaction order n_c. For generality, we allow the system to have $i = 1 \ldots N-1$ distinct forms of prefibrillar *cluster types*, with populations $f_i(n_c,t)$. The conversion rate constant for the process $f_i(n_c,t) \rightarrow f_{i+1}(n_c,t)$ is k_i^c, and only the Nth cluster type is able to elongate into fibrils at a rate defined by the elongation rate constant k_+. Reverse conversion and depolymerization reactions are not considered in this treatment, as under aggregation conditions they are slow in front of the forward rates.

This assembly process is described by a master equation for the number distributions $f_i(j,t)$ of aggregates of each type $i = 1 \ldots N$ with polymerization number j as follows:

$$\begin{aligned}
\dot{f}_1(j,t) &= \delta_{j,n_c}\left[k_n m(t)^{n_c} - k_1^c\, f_1(j,t)\right],\\
\dot{f}_l(j,t) &= \delta_{j,n_c}\left[k_{l-1}^c\, f_{l-1}(j,t) - k_l^c\, f_l(j,t)\right],\\
\dot{f}_N(j,t) &= 2\, m(t)\, k_+ [f_N(j-1,t) - f_N(j,t)]\\
&\quad + \delta_{j,n_c} k_{N-1}^c\, f_{N-1}(j,t),
\end{aligned} \tag{1}$$

where $2 \leqslant l \leqslant N-1$ and $m(t)$ gives the concentration of free monomer units available in solution for further polymerization. The Kronecker symbol is denoted by $\delta_{i,j}$, and the factor of 2 accounts for two free ends per fibril. Rather than

*Email address: tpjk2@cam.ac.uk

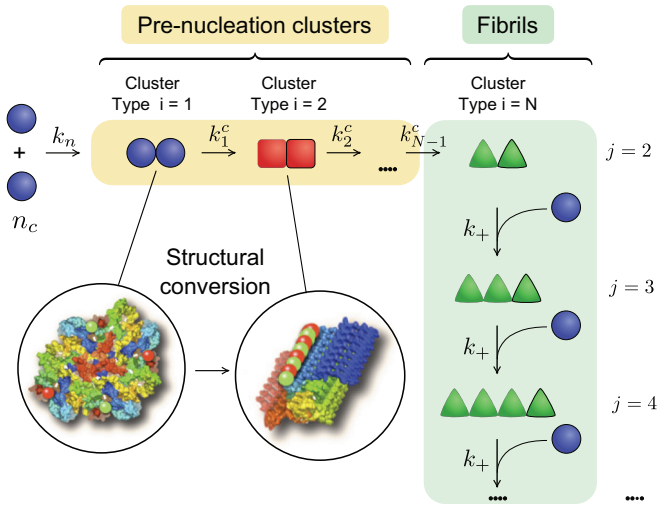


FIG. 1. (Color online) Schematic representation of a general primary nucleation conversion cascade, with $N-1$ different pre-fibrillar cluster types. The concentration of clusters of type i, with polymerization number j at time t, is denoted by the function $f_i(j,t)$. The primary nucleation reaction, of order n_c in free monomer, is described by the rate constant k_n; the subsequent conversion process $f_i(n_c,t) \to f_{i+1}(n_c,t)$ is described by the length-independent conversion rate constant k_i^c. Elongation of type-N clusters and fibrils is described by the rate constant k_+. Inset: Example of a particular physical realization of this general scheme: the conversion of an 8-mer of α-synuclein from a collapsed form to an ordered proteinase-K-resistant form [32] that proceeds to seed amyloid fibril formation.

distributions $f_i(j,t)$ of aggregates, the most commonly accessible observables are the total number or mass concentrations of species of each type [6,28,32]. A closed set of equations involving only these quantities is obtained through summation over the distributions $f_i(j,t)$, with appropriate treatment of the boundary terms [6], to yield the following:

$$\dot{P}_1(t) = k_n m(t)^{n_c} - k_1^c\, P_1(t); \quad M_1(t) = n_c\, P_1(t), \tag{2}$$

$$\dot{P}_l(t) = k_{l-1}^c\, P_{l-1}(t) - k_l^c\, P_l(t); \quad M_l(t) = n_c\, P_l(t), \tag{3}$$

$$\dot{P}_N(t) = k_{N-1}^c\, P_{N-1}(t), \tag{4}$$

$$\dot{M}_N(t) = 2\, m(t)\, k_+\, P_N(t) + k_{N-1}^c\, M_{N-1}(t), \tag{5}$$

where the number concentration $P_i(t) = \sum_j^\infty f_i(j,t)$ and the mass concentration $M_i(t) = \sum_j^\infty j\, f_i(j,t)$ are, respectively, the zeroth and first moments of the length distribution of the ith aggregate class.

III. ACCUMULATION OF SMALL SPECIES I

A central characteristic of a linear self-assembly system is whether the system primarily generates elongated fibrils or whether instead significant accumulation of smaller species occurs. The classical theory of nucleated polymerization [2] only allows for the formation of short fibrils, whereas the cascade nucleation process introduces the additional possibility of accumulation of prenucleation clusters. The accumulation of such species is determined by whether the majority of the initial system mass is sequestered into aggregates directly through the nucleation process, as described by the first term in Eq. (2) and resulting in small species, or is instead depleted through the growth of filaments, as described by the first term in Eq. (5) and resulting in elongated filaments. We can characterize this accumulation behavior by comparing the characteristic time scale for nucleation of aggregate mass, $\tau_n = (n_c\, k_n\, m_{\rm tot}^{n_c-1})^{-1}$ from Eq. (2), and the characteristic overall time scale for the assembly of elongated filaments, τ_a. In order to define τ_a, we note that the fibril mass $M_N(t)$ represents a central experimental observable [6,29] and its time dependence is sigmoidal for systems that form elongated fibrils; τ_a is thus evaluated as the time at which an inflection point is observed in $M_N(t)$ and will include contributions from conversion processes in addition to contributions from primary nucleation and elongation.

We begin by considering the early-time reaction kinetics [7] and examine first the behavior of a "filamentous" system, $\tau_a \ll \tau_n$, where significant accumulation of either short fibrils or intermediates does not occur. The effects of gradually increasing accumulation are then investigated, finally considering the limiting case of complete conversion of the system mass to small clusters.

IV. EARLY-TIME REACTION KINETICS

During the initial stages of the general aggregation reaction, only a small fraction of the available monomer units is incorporated into aggregates. For early reaction times $t \ll \tau_a$, the concentration of free monomer units is thus approximately constant at the initial value, $m(t) \approx m_{\rm tot}$, and Eqs. (2)–(5) can be linearized and readily analyzed through Laplace transformation. In the absence of seed material, the mass of filaments $M_N(t)$ is given approximately by

$$\frac{{}^0M_N(t)}{m_{\rm tot}} = \lambda^2 \left[\frac{1}{2}t^2 + \tau_c \left(\frac{\tau_c}{2} - t \right) - \sum_{i=1}^{N-1} \frac{\gamma_i\, e^{-k_i^c t} - \frac{1}{2}}{\left(k_i^c\right)^2} \right] + \tau_n^{-1} \left(t - \tau_c + \sum_{i=1}^{N-1} \frac{e^{-k_i^c t}}{k_i^c}\, \gamma_i \right), \tag{6}$$

where $\lambda = \sqrt{2\, k_+\, k_n m_{\rm tot}^{n_c}}$ is the effective rate constant controlling proliferation of elongated filaments in the absence of a cascade through intermediates (in which case $\tau_a = \lambda^{-1}$, denoting a general lower bound for τ_a) [7] and $\gamma_i = \prod_{j=1, j\neq i}^{N-1} k_j^c \cdot (k_j^c - k_i^c)^{-1}$ is a dimensionless weight factor. The presuperscript "0" denotes a linearized solution, and we identify $\tau_c = \sum_{i=1}^{N-1} (k_i^c)^{-1}$ as the characteristic conversion time controlling the cascade reaction. The second line in Eq. (6) represents the contribution from the accumulation of aggregate mass in small species. Provided that the extent of free monomer depletion remains small, $t \ll \tau_a$, and for times much less than the conversion time, $t \ll \tau_c$, the system described by Eq. (6) reduces to a polynomial of order dependent on N as

follows:

$$\frac{{}^0M_N(t)\big|_{t\to 0}}{m_{\rm tot}} = \tau_n^{-1}\frac{\kappa_c}{N!}\left(t^N - \frac{t^{N+1}}{N+1}\sum_{i=1}^{N-1}k_i^c\right) + \lambda^2\frac{\kappa_c}{N!}\left(\frac{t^{N+1}}{N+1}\right) + O(t^{N+2}), \quad (7)$$

where $\kappa_c = \prod_{i=1}^{N-1} k_i^c$ denotes the product of the conversion rate constants. For times much greater than the conversion time, such that $\tau_c \ll t \ll \tau_a$, the system is instead described by a quadratic form independent of N,

$$\frac{{}^0M_N(t)|_{t\to\infty}}{m_{\rm tot}} = \frac{1}{2}\lambda^2 t^2 + \left(\tau_n^{-1} - \lambda^2\tau_c\right)t + O(t^0). \quad (8)$$

These expressions, Eqs. (7) and (8), thus describe the early-time behavior of systems containing slow ($\tau_c \gg \tau_a$) and fast ($\tau_c \ll \tau_a$) cascades, respectively, where the characteristic conversion time is either much longer or much shorter than the time scale over which the overall fibril assembly reaction occurs; the *fast-* and *slow-conversion* early-time limits.

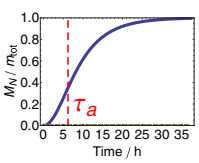
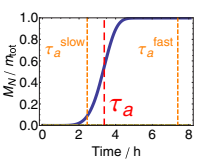
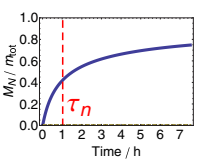
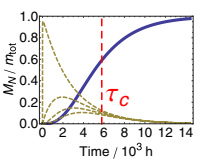
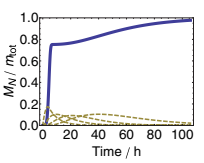
V. FAST AND SLOW CASCADES LEADING TO FIBRILS

The growth of type-N aggregates subsequent to the nucleation cascade results in the formation of highly elongated fibrils [32]. To study this behavior and its connections with the nucleation cascade, we now extend the range of validity of the early-time linearized solution Eq. (6) using a self-consistent argument and find an analytical expression for τ_a. When elongated fibrils are formed, the final elongation reaction sequesters the vast majority of the available free monomer units. Small aggregates thus account for a negligible proportion of the total system mass throughout the reaction. Accordingly, neglecting $M_i(t)$ for $2 \leqslant i \leqslant N-1$ reduces Eq. (5) to $\dot{M}_N(t) \approx 2\,m(t)\,k_+\,P_N(t)$, and conservation of mass gives $m_{\rm tot} \approx m(t) + M_N(t)$. These expressions can be combined and formally integrated to give $M_N(t) \approx 2\,k_+\,m_{\rm tot}\,e^{-2\,k_+\int_0^t P_N(t')\,dt'}\int_0^t P_N(t')e^{2\,k_+\int_0^{t'} P_N(t'')\,dt''}\,dt'$. This expression is self-consistent in (P_N, M_N) [6,7]; a first-order approximate solution ${}^1M_N(t)$ for the full time course thus can be found through the linearization $P_N(t) \approx {}^0P_N(t) \approx {}^0\dot{M}_N(t)/(2\,m_{\rm tot}\,k_+)$, remembering that ${}^0M_N(0) = 0$,

$$M_N(t) \approx {}^1M_N(t) = m_{\rm tot}\{1 - \exp[-{}^0M_N(t)/m_{\rm tot}]\}. \quad (9)$$

The time at which an inflection point is observed in ${}^1M_N(t)$ thus yields τ_a, leading to the analytical expression ${}^0\ddot{M}(\tau_a)m_{\rm tot} \approx ({}^0\dot{M}(\tau_a))^2$. As filamentous systems are characterized by a relatively slow nucleation reaction such that $\tau_n \gg \tau_a$ (and therefore $\tau_n^{-1} \ll \lambda$, as λ^{-1} is a lower bound for τ_a), the second line in Eq. (6) can be explicitly neglected to give an accurate implicit analytical expression for τ_a. Furthermore, by similarly neglecting terms containing τ_n^{-1}, the limiting cases of $\tau_c \gg \tau_a$ and $\tau_c \ll \tau_a$ yield the explicit analytical expressions $\tau_a^{\rm slow} = \sqrt[N+1]{N\,N!\lambda^{-2}\,\kappa_c^{-1}}$ and $\tau_a^{\rm fast} = \lambda^{-1} + \tau_c$ for slow and fast cascades, respectively, via Eqs. (7) and (8). The classical limit of $\tau_a = \lambda^{-1}$, with the well-known t^2 initial growth in $M_N(t)$ [2] from Eq. (8), is correctly recovered in the fast cascade *Oosawa limit* $\tau_c \to 0$. A more general *Oosawa-like* Regime I emerges, where the inclusion of a fast conversion reaction merely perturbs the overall aggregation kinetics of a filamentous system, $\{\tau_a \ll \tau_n, \tau_c < \lambda^{-1}\}$, as detailed in Table I. The early-time behavior in this regime is, hence, accurately captured by Eq. (8). By contrast, as τ_c increases such that $\{\tau_a \ll \tau_n,\ \tau_c \geqslant \lambda^{-1}\}$, the cascade time is no longer small relative to the overall aggregation kinetics, and fundamentally new behavior emerges. In this *slow-converting* regime, denoted II in Table I, the elongation of fibrils still dominates monomer

TABLE I. (Color online) Summary of different regimes that emerge for nucleation-conversion-polymerization systems. The time scale for assembly of elongated is fibrils denoted τ_a, with limiting values $\tau_a^{\rm slow} = \sqrt[N+1]{N\,N!\,\lambda^{-2}\,\kappa_c^{-1}}$ and $\tau_a^{\rm fast} = \lambda^{-1} + \tau_c$ for slow and fast conversion cascades, respectively. The time scale for assembly of elongated fibrils in the absence of a conversion cascade is given by $\lambda^{-1} = \sqrt{2\,k_n\,k_+\,m_{\rm tot}^{n_c-1}}$, and the time scale for conversion through a cascade of $N-1$ distinct prenucleation cluster species is given by $\tau_c = \sum_{i=1}^{N-1}(k_i^c)^{-1}$. The product of the conversion rate constants is denoted $\kappa_c = \prod_{i=1}^{N-1} k_i^c$, and the time scale for initial nucleation of clusters from monomer units is given by $\tau_n = (n_c\,k_n\,m_{\rm tot}^{n_c-1})^{-1}$. For each representative reaction time course, a numerical $M_N(t)$ solution representing fibril mass is shown in blue (solid line). Numerical $M_i(t)$ solutions for $1 \leqslant i \leqslant N-1$, representing the mass of each intermediate cluster type, are shown in dotted yellow; their mass is negligible in Regimes I, II, and III. Analytical predictions for the characteristic reaction time scale in Regimes I–IV are shown overlaid as dashed red vertical lines, via numerical solution in the case of Regime II of the algebraic equation ${}^0\ddot{M}(\tau_a)m_{\rm tot} \approx ({}^0\dot{M}(\tau_a))^2$, as discussed in the text, where ${}^0M_N(t)$ is obtained from the full first line of Eq. (6).

Regime	I Oosawa-like	II Slow-converting	III Nucleation-driven	IV Conversion-driven	V Two-stage
Representative reaction time course					
Elongated fibrils?	✓		×		(✓)
Conditions	$\tau_a \ll \tau_n$		$\tau_a \gg \tau_n$		$\tau_a \geqslant \tau_n$
	$\tau_c < \lambda^{-1}$	$\tau_c \geqslant \lambda^{-1}$	$\tau_c < \tau_n$	$\tau_c \geqslant \tau_n$	$\tau_c \gg \tau_n$
Characteristic time scale	$\tau_a \approx \tau_a^{\rm fast}$	$\tau_a^{\rm slow} < \tau_a \leqslant \tau_a^{\rm fast}$	τ_n	τ_c	Varies

depletion, but a pronounced lag phase emerges before $M_N(t)$ rises sharply with time, demonstrating sustained initial growth of higher order than t^2. Thus, the early-time behavior deviates qualitatively from the fast conversion limit Eq. (8) towards the slow conversion limit Eq. (7). The emergence of high-order growth is particularly interesting, as this behavior is traditionally attributed to secondary pathways [7].

VI. ACCUMULATION OF SMALL SPECIES II

In principle, a filamentous system in Regime II with a sufficiently slow cascade time and large number of cascade steps N can exhibit arbitrarily high-order initial growth, $M_N(t) \sim t^{N+1}$. But as this high-order slow conversion limit, $\{N \to \infty, \tau_c \to \infty\}$, is approached, significant accumulation of small aggregates becomes inevitable. This transition can be understood as an approach into a regime where $\tau_a \ll \tau_n$, such that primary nucleation becomes sufficiently fast to compete with mass sequestration from elongation; τ_n becomes smaller than the time scale for formation of elongated fibrils τ_a, and so the nucleation process begins to dominate the overall kinetics at all reaction times. Specifically, in the limit where $\tau_n \ll \lambda^{-1}$ (and therefore also $\tau_n \ll \tau_a$), early-time fibril mass formation is captured by the second line of Eq. (6) and, hence, by the terms containing τ_n^{-1} in the corresponding slow and fast conversion limits given by Eqs. (7) and (8).

Inspection of these resulting limits reveals two qualitatively distinct nonfilamentous regimes. In systems with fast primary nucleation and fast conversion, large numbers of short fibrils are produced, with minor accumulation of intermediates and with approximately linear initial growth of $M_N(t)$. Such a *nucleation-driven* Regime III with characteristic time τ_n, for $\{\lambda^{-1} \gg \tau_n, \tau_a \gg \tau_n, \tau_c < \tau_n\}$, also emerges from classical theory. However, it is rarely addressed in treatments of linearly aggregating systems as it does not lead to the formation of elongated fibrils. In addition, a *conversion-driven* Regime IV with characteristic time τ_c also emerges for the opposite conversion limit of $\{\lambda^{-1} \gg \tau_n, \tau_a \gg \tau_n, \tau_c \geqslant \tau_n\}$, where fast primary nucleation and slow conversion eventually lead to large numbers of short fibrils, via significant accumulation of intermediates and high-order initial polynomial growth of $M_N(t)$ with time. This regime, like Regime II, represents a generalization of the behavior described by the classical theory of nucleated polymerization. Finally, a semifilamentous transition regime also emerges between Regimes II and IV; this *two-stage* Regime V is described below.

VII. SEMIFILAMENTOUS GROWTH

Although the condition $\lambda^{-1} \gg \tau_n$ implies that $\tau_a \gg \tau_n$, the converse is not necessarily true when $\tau_c \geqslant \tau_n$; a slow conversion cascade ($\tau_c \geqslant \lambda^{-1}$, $\tau_c \gg \tau_n$) can lead to slow formation of fibril mass ($\tau_a \geqslant \tau_n$), even if the overall time scale for nucleation and elongation of fibril mass is fast ($\lambda^{-1} \leqslant \tau_n$). Under such conditions both lines of Eq. (6) contribute significantly to early-time fibril mass formation, thus producing a transition regime between filamentous and nonfilamentous growth. Such semifilamentous systems demonstrate an abrupt midreaction transition, from initial slow-conversion Regime II growth to conversion-driven Regime IV growth; a significant proportion of the available monomer units are sequestered within the conversion cascade such that the elongation reaction terminates prematurely. As this class of transition system exhibits clear qualitatively distinct behavior, we identify it as *two-stage* Regime V growth.

VIII. OVERALL BEHAVIOR

The various regimes and their properties are summarized in Table I. We explored the behavior of these aggregating systems by varying the rate constants by several orders of magnitude and solving the time dependence numerically (details of sampling procedure in Appendix). A comparison of a phase diagram generated in this manner with the analytical results derived in this work shows that the classification of aggregation behavior based on the dominant time scales is able to account well for the diversity in the aggregation behavior observed (Fig. 2).

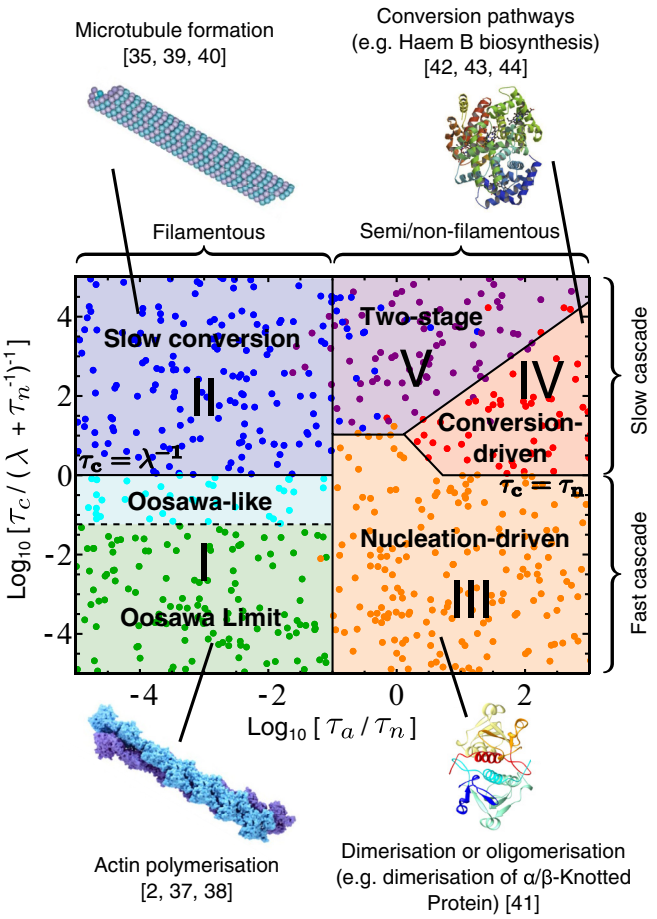


FIG. 2. (Color online) Phase diagram illustrating the relationship between the different growth regimes predicted for nucleation-conversion-polymerization reactions, based on numerical simulations of sampled systems (described in Appendix). The horizontal axis gives the ratio of the characteristic time scale for the assembly of elongated filaments τ_a to the primary nucleation time scale τ_n. The vertical axis compares the characteristic conversion time τ_c to a measure of the nucleation-elongation time scale, represented by the expression $(\lambda + \tau_n^{-1})^{-1}$ which interpolates correctly between the relevant limiting expressions, λ^{-1} as $k_+/(k_n m_{\mathrm{tot}}^{n_c-2}) \to \infty$ and τ_n as $k_+/(k_n m_{\mathrm{tot}}^{n_c-2}) \to 0$. Biologically relevant examples of Regimes I–IV are also shown. Images modified and reproduced from Ref. [38] (actin), Ref. [40] (tubulin), Ref. [41] (α/β-knotted protein), and Ref. [44] (hemoglobin structure).

IX. HIGH-ORDER GROWTH AND SECONDARY PATHWAYS

The emergence of high-order growth seen in Regime II is of particular interest for understanding aggregation phenomena that are commonly observed for proteins *in vitro* [14,28,45]. Experimental observations of steeper-than-quadratic growth with time have often been interpreted as indicative of secondary pathways, such as fibril fragmentation and surface-catalyzed secondary nucleation [1,7]. The present work shows that such behavior can also emerge, under appropriate conditions, from a primary nucleation process. The present results also outline, however, a strategy to discriminate between primary and secondary nucleation. For example, the addition of preformed seed fibrils to a Regime II cascade nucleation system allows the immediate sequestration of mass via elongation, bypassing the slow conversions steps and thus significantly reducing the order of initial growth in $M_N(t)$. In contrast, for a system with secondary pathways, seeding will reduce the observed reaction lag times but will still produce high-order growth [29]. Variation of the total monomer concentration m_{tot} will also produce different effects in each case; as shown for a Regime II cascade nucleation system, the characteristic time τ_a is expected to scale with m_{tot}^{α}, where α lies in the range $-n_c/2 \leqslant \alpha \leqslant -n_c/(N+1)$—a clear generalization of the classical result $\alpha = -n_c/2$. A qualitative transition may also be observed into a different growth regime. In contrast, fragmentation and secondary nucleation predict characteristic times that scale with $\alpha = -1/2$ and $\alpha = -(n_2+1)/2$, respectively, where n_2 denotes the critical nucleus size for secondary nucleation [7]. The scaling of characteristic times thus further facilitates discrimination among different pathways, a crucial step in establishing the molecular mechanisms underlying protein aggregation phenomena in nature.

The analysis presented here quantitatively delimits and characterises five qualitatively different growth regimes for nucleation-conversion-elongation self-assembly reactions, recovering and generalizing results obtained from the classical theory of nucleated polymerization. Furthermore, analytical expressions are found for the characteristic reaction times in each case, and a range of detailed analytical expressions are presented for the purposes of interrogating experimental results.

ACKNOWLEDGMENTS

The authors thank the Schiff Foundation, the Newman Foundation, and the BBSRC for their generous support.

APPENDIX: CONSTRUCTING THE PHASE DIAGRAM

1. Sampling the parameter space

The phase diagram (Fig. 2) demonstrating the position of each regime in the nucleation-conversion-polymerization parameter space was generated by numerical random sampling. The diagram coordinates $\{x,y\}$ were sampled log-uniformly such that $x = 10^X$ and $y = 10^Y$, where both X and Y are independent random variables uniformly distributed in the ranges $[-5,5]$ inclusive. Other system parameters were sampled from physically and biologically relevant ranges as follows:

N: Random integer, sampled uniformly in the range $[2,15]$ inclusive,

n_c: Random integer, sampled uniformly in the range $[2,6]$ inclusive,

$m_{\text{tot}} = 10^G$ mol dm^{-3}, where G is a random real number sampled uniformly in the range $[-8,-2]$,

$k_i^c = 10^{H_i}$ s^{-1} for $1 \leqslant i \leqslant N-1$, where $H_1, H_2, \ldots, H_{N-1}$ are independent random real numbers, sampled uniformly in the range $[-5,5]$.

The rate constants k_+ and k_n were determined by expressing the axes of the phase diagram in terms of other system parameters: $x = \tau_a/\tau_n \approx (\lambda^{-1} + \tau_c)/\tau_n$ and $y = \tau_c/(\lambda + \tau_n^{-1})^{-1}$, where $\lambda = \sqrt{2\,k_+\,k_n\,m_{\text{tot}}^{n_c}}$, $\tau_n = (n_c\,k_n\,m_{\text{tot}}^{n_c-1})^{-1}$, and $\tau_c = \sum_{i=1}^{N-1}(k_i^c)^{-1}$. The bulk rate constants λ and τ_n thus can be expressed in terms of the coordinates $\{x,y\}$,

$$\begin{pmatrix} \lambda\,\tau_c \\ \tau_n\,\tau_c^{-1} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1-x+y \\ \frac{1+x+y}{x\,y} \end{pmatrix} \pm \frac{1}{2}\sqrt{(-1-x+y)^2+4\,y}\begin{pmatrix} 1 \\ (x\,y)^{-1} \end{pmatrix}.$$

By taking the additive solution above, λ and τ_n are always positive, and the remaining microscopic rate constants $k_+ = n_c\,\tau_n\,\lambda^2/(2\,m_{\text{tot}})$ and $k_n = (n_c\,\tau_n\,m_{\text{tot}}^{n_c-1})^{-1}$ can be determined.

2. Numerical simulation of systems

Once established, the system parameters were introduced into the moment equations Eqs. (2)–(5), and the system was numerically integrated using the NumPY and SciPY packages in PYTHON 2.7, with the boundary condition that at $t = 0$ the system consists solely of free monomer units. For each system, the values of $M_N(t)$ and $P_N(t)$ were determined at 100 000 linearly evenly-spaced points in time from $t = 0$ to $t = t_{\text{max}}$, where t_{max} is defined as $10 \times \tau_c^{\text{fast}}$ or, equivalently, $10\,(\lambda^{-1} + \tau_c)$; this time range appropriately captured the reaction in the vast majority of cases.

3. Determining and representing different growth regimes

The growth regime for each system was determined by the following protocol:

(1) Ensure that there are no numerical overflows or obvious errors in the numerical integration [for example, ensure that $\dot{M}_N(t)$ remains positive at all times]. Otherwise, discard system.

(2) Ensure that by $t = t_{\text{max}}$, the reaction is largely complete: either $M_N(t_{\text{max}})/m_{\text{tot}} \geqslant 0.95$, or, alternatively, $M_N(t_{\text{max}})/m_{\text{tot}} \geqslant 0.5$ *and* $[M_N(t_{\text{max}}) - M_N(0.9\,t_{\text{max}})]/m_{\text{tot}} \leqslant 0.1$, such that $M_N(t)/m_{\text{tot}}$ is changing very slowly. Under either set of conditions, the average fibril length $M_N(t)/P_N(t)$ will not change by more than $\sim$5% beyond $t = t_{\text{max}}$. If conditions are not met, discard system (or, alternatively, extend integration time until conditions are met; here offending systems were discarded).

(3) If more than one inflection point is apparent in $M_N(t)$, record the value of $v = M_N(t)/m_{\rm tot}$ at the second inflection point. If $0.02 \leqslant v \leqslant 0.98$, assign *Regime V* two-stage growth. The system is semifilamentous.

(4) If less than two inflection points are visible, or if the specified conditions for v are not met, evaluate the average fibril length l numerically at $t_{\rm max}$, given by $l = M_N(t_{\rm max})/P_N(t_{\rm max})$. If $l > 20\,n_c$, the system is *filamentous*. Otherwise, it is nonfilamentous.

(5) If the system is filamentous, evaluate $\theta = \tau_c/\lambda^{-1}$. If $\theta < 0.05$, assign *Regime Ia* Oosawa limit growth. If, instead, $0.05 \leqslant \theta < 1$, assign *Regime Ib* Oosawa-like growth. Otherwise, if $\theta \geqslant 1$, assign *Regime II* slow-conversion growth.

(6) If instead the system is nonfilamentous, evaluate $\phi = \tau_c/\tau_n$. If $\phi < 1$, assign *Regime III* nucleation-driven growth. Otherwise, if $\phi \geqslant 1$, assign *Regime IV* conversion-driven growth.

Each sampled system was then represented as a colored point at position $\{x,y\}$ on the phase diagram, with the x coordinate evaluated precisely as τ_a/τ_n by solving the algebraic equation ${}^0\ddot{M}_N(\tau_a)m_{\rm tot} = ({}^0\dot{M}_N(\tau_a))^2$ numerically using the first line of Eq. (6), as described in the main text. Both dimensionless axes are represented as base-10 logarithms, with colors chosen to correspond with different growth regimes as follows: Ia, green (Oosawa limit); Ib, cyan (Oosawa-like); II, blue (slow-conversion); III, orange (nucleation-driven); IV, red (conversion-driven); and V, purple (two-stage).

Simple regime boundaries were then drawn to represent the resulting distribution of points, and the different regions on the phase diagram were shaded and labeled accordingly, with the diagram cropped to highlight all interesting behavior.

[1] F. Oosawa and M. Kasai, J. Mol. Biol. **4**, 10 (1962).
[2] F. Oosawa and S. Asakura, *Thermodynamics of the Polymerization of Protein* (Academic Press, New York, 1975).
[3] H. X. Zhou and F. A. Ferrone, Biophys. J. **58**, 695 (1990).
[4] F. Ferrone, J. Hofrichter, and W. A. Eaton, J. Mol. Biol. **183**, 611 (1985).
[5] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman, PLoS Biol. **2**, e321 (2004).
[6] 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).
[7] S. I. A. Cohen, M. Venduscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, and T. P. J. Knowles, J. Chem. Phys. **135**, 065105 (2011).
[8] S. I. A. Cohen, M. Venduscolo, C. M. Dobson, and T. P. J. Knowles, J. Chem. Phys. **135**, 065106 (2011).
[9] S. I. A. Cohen, M. Venduscolo, C. M. Dobson, and T. P. J. Knowles, J. Chem. Phys. **135**, 065107 (2011).
[10] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, Int. J. Mol. **12**, 5844 (2011).
[11] L. S. Tobacman and E. D. Korn, J. Biol. Chem. **258**(5), 3207 (1983).
[12] C. M. Dobson, Trends Biochem. Sci. **24**, 329 (1999).
[13] F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).
[14] C. M. Dobson, Nature **426**, 884 (2003).
[15] F. E. Cohen and J. W. Kelly, Nature **426**, 905 (2003).
[16] D. J. Selkoe, Nature **426**, 900 (2003).
[17] P. T. Lansbury and H. A. Lashuel, Nature **443**, 774 (2006).
[18] M. B. Pepys, Phil. Trans. R. Soc. Lond. B **356**, 203 (2001).
[19] J. Hardy and D. J. Selkoe, Science **297**, 353 (2002).
[20] J. C. Sacchettini and J. W. Kelly, Nat. Rev. Drug Discov. **1**, 267 (2002).
[21] S. B. Prusiner, Science **252**, 1515 (1991).
[22] J. H. Come, P. E. Fraser, and P. T. Lansbury, Proc. Natl. Acad. Sci. USA **90**, 5959 (1993).
[23] A. Aguzzi and C. Haass, Science **302**, 814 (2003).
[24] J. Falsig, K. P. R. Nilsson, T. P. J. Knowles, and A. Aguzzi, HFSP J. **2**, 332 (2008).
[25] A. Aguzzi and T. O'Connor, Nat. Rev. Drug Discov. **9**, 237 (2010).
[26] J. Hofrichter, P. D. Ross, and W. A. Eaton, Proc. Natl. Acad. Sci. USA **71**, 4864 (1974).
[27] J. Hofrichter, J. Mol. Biol. **189**, 553 (1986).
[28] F. Ferrone, Methods Enzymol. **309**, 256 (1999).
[29] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, J. Mol. Biol. **421**, 160 (2012).
[30] M. F. Bishop and F. A. Ferrone, Biophys. J. **46**, 631 (1984).
[31] A. K. Buell, J. R. Blundell, C. M. Dobson, M. E. Welland, E. M. Terentjev, and T. P. J. Knowles, Phys. Rev. Lett. **104**, 228101 (2010).
[32] N. Cremades, S. I. A. Cohen, E. Deas, A. Y. Abramov, A. Y. Chen, A. Orte, M. Sandal, R. W. Clarke, P. Dunne, F. A. Aprile, C. W. Bertoncini, N. W. Wood, T. P. J. Knowles, C. M. Dobson, and D. Klenerman, Cell **149**, 1048 (2012).
[33] J. Lee, E. K. Culyba, E. T. Powers, and J. W. Kelly, Nat. Chem. Biol. **7**, 602 (2011).
[34] A. Vitalis and R. V. Pappu, Biophys. Chem. **159**, 14 (2011).
[35] H. Flyvbjerg and S. Leibler, Proc. Natl. Acad. Sci. USA **93**, 5975 (1996).
[36] F. A. Ferrone, Methods Enzymol. **412**, 285 (2006).
[37] E. Nishida and H. Sakai, J. Biochem. **93**(4), 1011 (1983).
[38] U. S. National Library of Medicine (2013), http://ghr.nlm.nih.gov/handbook/illustrations/actin.
[39] W. A. Voter and H. P. Erickson, J. Biol. Chem. **259**(16), 10430 (1984).
[40] C. Conde and A. Cáceres, Nat. Rev. Neurosci. **10**, 319 (2009).
[41] A. L. Mallam and S. E. Jackson, Structure **15**, 111 (2007).
[42] C. Rimington, Brit. Med. J. **2**(4986), 189 (1956).
[43] S. Sassa and A. Kappas, J. Clin. Invest. **71**, 625 (1983).
[44] M. A. Schumacher, E. E. Zheleznova, K. S. Poundstone, R. Kluger, R. T. Jones, and R. G. Brennan, Proc. Natl. Acad. Sci. USA **94**, 7841 (1997).
[45] T. P. Knowles, A. W. Fitzpatrick, S. Meehan, H. R. Mott, M. Vendruscolo, C. M. Dobson, and M. E. Welland, Science **318**, 1900 (2007).