ARTICLE

# Dissecting the Kinetic Process of Amyloid Fiber Formation through Asymptotic Analysis

Liu Hong,[†,‡] Xianghong Qi,[†] and Yang Zhang[*,†,§]

[†]Center for Computational Medicine and Bioinformatics and [§]Department of Biological Chemistry, University of Michigan, 100 Washtenaw Avenue, Ann Arbor, Michigan 48109, United States

[‡]Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University, Beijing 100084, P. R. China

Ⓢ *Supporting Information*

ABSTRACT: Amyloids are insoluble fibrous protein aggregates which, when abnormally accumulated in the body, can result in amyloidosis and various neurodegenerative diseases. In this work, we describe a new approach to the asymptotic solution of the master equation of amyloid fiber aggregations. It is found that four distinct and successive stages (lag phase, exponential growth phase, breaking phase, and static phase) dominate the fiber formation process. On the basis of the distinctive power-law dependence of the half-time and apparent growth rate of the fiber formation on the initial protein concentration, we propose a novel classification for amyloid proteins theoretically.
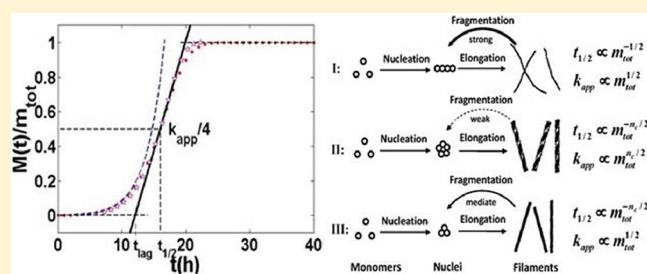
INTRODUCTION

Abnormal accumulation of amyloid fiber in organs is closely associated with various neurodegenerative ailments, including Alzheimer's and Huntington's diseases.[1−3] Uncovering the mechanism of amyloid fiber formation will provide an important step forward in identifying the onset of the amyloid-related diseases and in developing new antifibrotic treatments.[4,5] Throughout the past decade, a number of processes have been experimentally identified in the amyloid fiber formation, including protein unfolding and refolding,[6−8] on- and off-pathway competition,[9] homo- or heterogeneous nucleation,[10,11] autocatalytic surface growth,[12] elongation,[13] merging,[14] fragmentation,[15,16] branching,[17] thickening,[14,18,19] etc. Despite extensive studies, a general mechanism to account for the various amyloid aggregation processes still remains to be elucidated.

Protein unfolding and refolding are prerequisite steps for amyloid fiber formation. Prior to fibrillation, the tertiary structure of amyloid proteins is either globular or intrinsically disordered with the secondary structure of the precursor proteins changing in the α-helix and β-sheet contents. After fibrillation, however, all protein fibrils share a unified morphology—twisted and ropelike with a filamentous substructure where continuous β-sheets are formed with the β-strands running perpendicular to the vertical axis of the fibrils.[20] The initial formation of a critical nucleus is believed to be a central step for the onset of amyloid fibrillation,[21−25] but the actual function of the nucleation process is yet to be fully understood. One difficulty is the complex oligomers with different size formed simultaneously in the system which engenders serious difficulties in the identification of the critical nucleation events. In addition, the traditional monomer-concerned nature of the primary nucleation cannot provide a reasonable answer to current agitating studies, which have in fact highlighted the necessity of a fibril-dependent secondary nucleation.[15,16] Fragmentation events may serve as a key process for monomer-independent secondary nucleation, helping to explain the exponential growth elements of the sigmoidal curve.[26−29] Another crucial aspect in amyloid fiber formation is elongation, a process by which small oligomers or protofibrils grow into the characteristic long mature fibrils upon the linear addition of monomers to their terminal ends. The importance of other processes (including branching, merging, thickening etc.) may vary from protein to protein. For instance, glucagon fibrils have been observed branching continuously, whereas the fibril growth of Aβ(1−40) remains strictly linear.[17]

Another difficulty in the treatment of amyloid fibril diseases is the lack of a clear and complete physical description encompassing the entire process of amyloid fiber formation. Although some empirically defined characteristic stages have been proposed, such as the lag and exponential growth phases, there is a lack of mathematical verifications supported by quantitative data. For example, what are the quantitative differences between different aggregation phases? Is there a general mechanism governing different processes? To answer these questions, a quantitative description and classification of the amyloid fibrillation process is demanded.
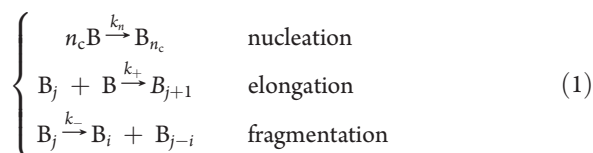
Knowles et al. recently presented a chemical master-equation based model, which involves three steps of amyloid fiber formation:[30] primary nucleation, elongation, and fragmentation. The modeling results were found in agreement with several aspects of experimental data, which partly validated the general assumptions regarding to the mechanisms of fibril formation. However, a qualitative explanation of the fibrillation process is not straightforward from the approximate solutions obtained via the fixed-point analysis by Knowles et al. Meanwhile, their solutions are not applicable to the static state.

In this work, we aim at presenting a complete and quantitative description for the formation of amyloid fibers. We first explore the Knowles et al. model through asymptotic analysis, an approximation method based on the mathematical analysis of partial differential equations within different time scales, which thereby provides a quantitative dissection of the kinetic process of amyloid fiber formation. Second, on the basis of the analytical solutions of the half-time and apparent growth rate of fiber formation, in particular their distinctive power-law dependence on the initial soluble protein concentration, we propose a new classification of amyloid proteins.

■ MATERIALS AND METHODS

A. Basic Equations. We consider three fundamental processes of fibrous aggregations which can be expressed through following chemical reactions:

$$
\begin{cases}
n_c B \xrightarrow{k_n} B_{n_c} & \text{nucleation} \\
B_j + B \xrightarrow{k_+} B_{j+1} & \text{elongation} \\
B_j \xrightarrow{k_-} B_i + B_{j-i} & \text{fragmentation}
\end{cases}
\tag{1}
$$

where B stands for a monomer and B_j for filaments with length j. We have assumed $1 \leq i < j$, $j \geq n_c$, where $n_c \geq 2$ is the minimum size of filaments. Here for modeling facility, we have only included the forward reactions, but no corresponding backward reactions. Though this may be a reasonable approximation for the fiber growth phase, it will lead to some unrealistic predictions when we are dealing with a genuine thermodynamic equilibrium. Besides, we have also assumed there is no preference position for fiber breakage, and the fiber breaking rate is independent of fiber length. Yet both experiments and theoretical analysis by Hill[49] showed that filaments can be more easily broken in the middle, and short filaments are much harder to break than long ones. These limitations will be amended in our future studies.

If we define $f(t,j)$ as the concentration of filaments in the system with length j at time t and $m(t)$ as the concentration of monomers, we can use the master equations to describe the time evolution of $f(t,j)$ in the processes of eq 1:[30]

$$
\frac{\partial f(t,j)}{\partial t} = 2m(t)k_+[f(t,j-1) - f(t,j)] - k_-(j-1)f(t,j)
$$

$$
+ 2k_- \sum_{i=j+1}^{\infty} f(t,i) + k_n m(t)^{n_c} \delta_{j,n_c}
\tag{2}
$$

The first term on the right side of above equation is related with fiber elongation, the second and third one with fiber fragmentation,[52] and the last one stands for the nucleation process. The summations

of eq 2 will result in the evolution master equations of the total filament number and mass concentrations:

$$
\begin{cases}
\dfrac{dP(t)}{dt} = k_-[M(t) - (2n_c - 1)P(t)] + k_n m(t)^{n_c} \\[2mm]
\dfrac{dM(t)}{dt} = [2k_+ m(t) - n_c(n_c - 1)k_-]P(t) + n_c k_n m(t)^{n_c}
\end{cases}
\tag{3}
$$

where $P(t) = \sum_{j=n_c}^{\infty} f(t,j)$ is the total number of filaments in the system, $M(t) = \sum_{j=n_c}^{\infty} j \times f(t,j)$ is the mass concentration, k_n, k_+, and k_- are the reaction rates for primary nucleation, elongation, and fragmentation, respectively, as defined in eq 1. We also have $m(t) = m_{tot} - M(t)$ according to the mass conservation law, where m_{tot} is the total protein concentration, given that the contributions from oligomers with size $1 < n < n_c$ are negligible.

B. Empirical Formula for Amyloid Fibrillation. Experimental values for the apparent fiber growth rate k_{app} and the half-time of fiber formation $t_{1/2}$ are usually determined through the following empirical equation

$$
F(t) = F(0) + \frac{A}{1 + \exp[-k_{app}(t - t_{1/2})]}
\tag{4}
$$

which can best fit the measured time evolutionary curves for the ThT fluorescence intensity during amyloid fibrillation.[31] Here, $F(t)$ is the ThT fluorescence intensity at time t, $F(0)$ is the background ThT fluorescence intensity at the starting time, and A is a constant.

C. Determination of Model Parameters. The unknown reaction rates of elongation, fragmentation and nucleation (k_+, k_-, k_n) in eq 1 can be determined by

$$
\begin{cases}
k_+ = \sqrt{2\alpha n_c(n_c - 1)}\kappa/(2m_{tot}) \\
k_- = \kappa/\sqrt{2\alpha n_c(n_c - 1)} \\
k_n = \kappa\sigma/(\sqrt{2\alpha n_c(n_c - 1)}m_{tot}^{n_c - 1})
\end{cases}
\tag{5}
$$

where the intermediate variables κ, α, and σ are expected to be correlated with the half-time of fiber formation $t_{1/2}$, the apparent fiber growth rate k_{app}, and the equilibrium mass fraction of monomers χ through

$$
\begin{cases}
\alpha = 1/(2\chi) \\
\kappa = k_{app}/\sqrt{2} \\
\sigma = \exp(-k_{app}t_{1/2}/\sqrt{2})
\end{cases}
\tag{6}
$$

These relationships[50] can be summarized into an automated procedure for the model parameter determination, i.e., $(t_{1/2}, k_{app}, \chi) \rightarrow (\alpha, \kappa, \sigma) \rightarrow (k_+, k_-, k_n)$. As shown in Figure S2 (Supporting Information), the estimated model parameters by this procedure and the experimental data are highly correlated with a Pearson correlation coefficient ~0.95.

■ RESULTS

The typical evolutionary of $P(t)$ and $M(t)$ based on eq 3 is shown in Figure 1. We can identify three distinguishable phases from the sigmoidal curve of $M(t)$: an initial lag phase (nucleation dominated), an intermediate stage with rapid increase (the exponential growth phase), followed by a final plateau phase. Compared to the time evolution of $P(t)$, the plateau phase can be further subdivided into two different parts: the first with constant
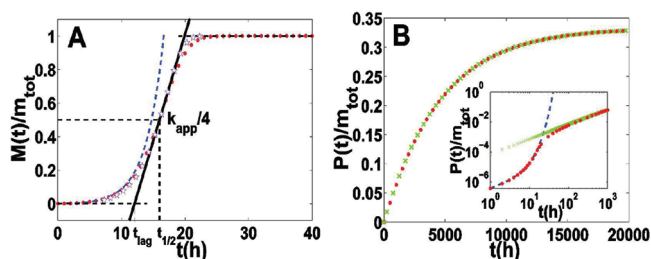
6612

dx.doi.org/10.1021/jp205702u |J. Phys. Chem. B 2012, 116, 6611–6617

Figure 1. Comparison of asymptotic solutions and numerical solutions of eq 3. The red dots represent the numerical solutions of eq 3, the dashed blue line represents asymptotic solutions for the nucleation phase obtained through eq 7, the pink stars show the asymptotic solution for the exponential growth phase obtained through eq 10, and the green crosses show the asymptotic solution for the breaking phase obtained through eq 12. Here we set $n_c = 2$, $k_+ = 5 \times 10^4 \, \mathrm{M^{-1} \, s^{-1}}$, $k_- = 2 \times 10^{-8}$ $\mathrm{s^{-1}}$, $k_n = 2 \times 10^{-5} \, \mathrm{M^{-1} \, s^{-1}}$, $m_{tot} = 5 \times 10^{-6}$ M.

$M(t)$ and increasing $P(t)$ (the breaking phase), and the second with constant $P(t)$ and $M(t)$ (the final static phase). Therefore, the whole process of amyloid fiber formation can be formally divided into four successive stages: an initial lag phase, an exponential growth phase, a continuously breaking phase, and a final static phase. Each stage is characterized by its particular time scale, along with the number and mass concentrations of filaments.

In the following sections, we will further validate these four stages through detailed mathematical analysis. We will focus on the half-time and apparent growth rate of fiber formation, which have been widely used in the empirical interpretation of experimental data for amyloid fibrillation. The scaling relationships among the half-time of fiber formation, the apparent fiber growth rate, and the initial protein concentration are fully explored, which results in a tentative classification of amyloid proteins based on the differences in their dominant nucleation mechanisms.

A. Lag Phase. For the experiments on amyloid fiber formation without seeding, we can observe a long initial time period (referred to as "lag-time"), in which most amyloid proteins remain in the soluble monomer form and very few oligomers or filaments can be detected.[32] Within this lag-time, the combination process of multiple monomers into stable or unstable nuclei dominates. Because $m(t) \cong m_{tot}$, we can neglect the fragmentation terms involving $P(t)$, and replace $m(t)$ by m_{tot} in eq 3, i.e.

$$\begin{cases} \dfrac{dP(t)}{dt} = k_- M(t) + k_n m_{tot}^{n_c} \\ \dfrac{dM(t)}{dt} = 2k_+ m_{tot} P(t) + n_c k_n m_{tot}^{n_c} \end{cases} \tag{7}$$

which gives

$$\begin{cases} P(t) = (C_+ \kappa e^{\kappa t} - C_- \kappa e^{-\kappa t} - n_c k_- \sigma m_{tot})/(2k_+ m_{tot}) \\ M(t) = C_+ e^{\kappa t} + C_- e^{-\kappa t} - \sigma m_{tot} \end{cases} \tag{8}$$

where

$$\begin{cases} \sigma = k_n m_{tot}^{n_c - 1}/k_- \\ \kappa = \sqrt{2k_+ k_- m_{tot}} \\ C_\pm = \dfrac{1}{2}\{M(0) + \sigma m_{tot} \pm [n_c k_- \sigma m_{tot} + 2k_+ m_{tot} P(0)]/\kappa\} \end{cases} \tag{9}$$

From this solution, we can have $\dot{M}(0) \approx 2m_{tot}k_+P(0) = 2m_{tot}k_+M(0)/L_0$ under the condition of "seeding" where L_0 is the initial average length of filaments. This indicates that the initial growth rate of a seeded fibrillation reaction is proportional to the initial concentrations of soluble proteins and seeds.

B. Exponential Growth Phase. After certain accumulation of fiber nuclei $(P(t) > n_c k_n m_{tot}^{n_c - 1}/k_+)$, the mass concentration of filaments $M(t)$ enters a phase of exponential growth. The major contribution to this rapid increase comes from two aspects: one is elongation, which quickly promotes short filaments into longer ones via monomer addition; another is fragmentation, which provides new seeds through the spontaneous breaking of long filaments into shorter ones. The former aspect provides a polynomial increase in the mass concentration of filaments, whereas the latter engenders the exponential growth.[33] As the change in the mass concentration of filaments $M(t)$ is usually much faster than the number concentration of filaments $P(t)$ in current stage (Figure 1), we can adopt the method of variable separation, and assume that the slow variable $P(t)$ still obeys the equation $P(t) = \sigma[(\kappa + n_c k_-)e^{\kappa t} - (\kappa - n_c k_-)e^{-\kappa t} - 2n_c k_-]/(4k_+)$ obtained in the nucleation stage, whereas the fast variable $M(t)$ satisfies

$$\frac{dM(t)}{dt} = 2k_+ m(t)\, P(t) \tag{10}$$

by only keeping the elongation term. The solution is given as $M(t) = m_{tot}\{1 - \exp[-\sigma(1 + n_c k_-/\kappa)e^{\kappa t}/2 - \sigma(1 - n_c k_-/\kappa)e^{-\kappa t}/2 + n_c k_- \sigma t + \sigma]\}$.

In the current study, two quantities are most interested: one is the half-time of fiber formation, which refers to the time point when mass concentration reaches 50% of the final value, i.e., $M(t_{1/2})/M(\infty) = 50\%$. The other is the apparent fiber growth rate, given by four times of the slope of sigmoid curve at the half-time normalized by initial protein concentration,[32] i.e.

$$\begin{cases} k_{app} \sim \kappa & t_{1/2} \sim \ln(1/\sigma)/\kappa & \sigma \ll 1 \\ k_{app} \sim \sqrt{\sigma}\kappa & t_{1/2} \sim 1/(\sqrt{\sigma}\kappa) & \sigma \gg 1 \end{cases} \tag{11}$$

These equations can be further simplified as $k_{app} \propto m_{tot}^{1/2}$, $t_{1/2} \propto m_{tot}^{-1/2}$ when $\sigma \ll 1$; $k_{app} \propto m_{tot}^{n_c/2}$, $t_{1/2} \propto m_{tot}^{n_c/2}$ when $\sigma \gg 1$, which offers a direct theoretical basis for our later classification of amyloid proteins in section E.

C. Breaking Phase. Following the exponential growth phase, the mass concentration of filaments reaches almost constant with $M(t) \approx m_{tot}[1 - 1/(2\alpha)]$, where $\alpha = k_+ m_{tot}/[k_- n_c(n_c - 1)]$. The number concentration of filaments increases until $P(t) \approx M(t)/(2n_c - 1)$. As a result, the average length of filaments $M(t)/P(t)$ decreases continuously from the maximum value $(2k_+ m_{tot}/k_-)^{1/2}$ in the exponential growth phase to the final value $2n_c - 1$. The governing equation for the number concentration of filaments in this stage can be approximated as

$$\frac{dP(t)}{dt} = k_-[M(t) - (2n_c - 1)P(t)] \tag{12}$$

which gives $P(t) = m_{tot}[1 - 1/(2\alpha)][1 - e^{-(2n_c-1)k_- t}]/(2n_c - 1)$ (Figure 1). In fact, this equation is consistent with the numerical solution of $P(t)$ in nearly the whole time range.

From eq 12, the time scale of the breaking phase is given by $t_b \approx 1/[(2n_c - 1)k_-]$, which is usually much larger than the half-time of fiber formation. For instance, the parameters that best fit the kinetic curve of WW domain fibrillation[34] at the initial
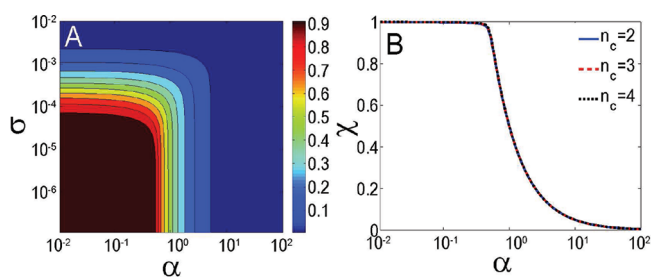
Figure 2. (A) Contour plot of the static mass fraction of monomers χ as a function of parameters α and σ with $n_c = 2$. (B) Dependence of the static mass fraction of monomers χ on parameter α. All of the curves with various nucleus size n_c converge into the same line.

concentration $m_{tot} = 50\,\mu M$, are estimated to be $n_c = 2$, $k_+ = 1.94 \times 10^4\,M^{-1}\,s^{-1}$, $k_- = 9.7 \times 10^{-9}\,s^{-1}$, $k_n = 1.2 \times 10^{-7}\,M^{-1}\,s^{-1}$. We obtained $t_{1/2} \approx 15.1$ h and $t_b \approx 1.1$ years, showing that the time scale of breaking phase is 600 times longer than the half-time of fiber formation. Although there are still lots of controversies, we guess such a large time scale for the fiber breaking phase may correlate with the long latent period of amyloid-related diseases. It has been suggested by many experimental studies that small oligomers and protofilaments are more toxic to cells than larger mature filaments.[16,34] Due to the influence of the fiber breaking process, more small oligomers will bring increasing damage to cells, which eventually leads to clinically observed symptoms.[2,3]

D. Static Phase. Finally, the system will reach a static state with constant values of $P(t)$ and $M(t)$. Letting $t \rightarrow \infty$, the asymptotic behaviors of eq 3 become

$$\begin{cases} 0 = k_-[m_{tot} - m(\infty) - (2n_c - 1)P(\infty)] + k_n m(\infty)^{n_c} \\ 0 = 2[m(\infty)k_+ - n_c(n_c - 1)k_-/2]P(\infty) + n_c k_n m(\infty)^{n_c} \end{cases}$$

$$(13)$$

where $m(\infty)$ and $P(\infty)$ indicate the static values of $m(t)$ and $P(t)$ when $t \rightarrow \infty$.

If we define $\chi = m(\infty)/m_{tot}$ as the static mass fraction of monomers in the system, eq 13 can be written as a closed algebraic function of χ, i.e.

$$2\alpha\sigma\chi^{n_c + 1} + \frac{n_c}{n_c - 1}\sigma\chi^{n_c} - 2\alpha\chi^2 + (2\alpha + 1)\chi - 1 = 0$$

$$(14)$$

The typical dependence of the static mass fraction of monomers χ on parameters α and σ is illustrated in Figure 2A, with $n_c = 2$. For larger values of n_c, the basic pattern of the contour plot of χ does not change, except for a barely perceptible shift toward a larger value of n_c. Although eq 14 involves three parameters (α, σ, n_c), we demonstrate that χ depends mainly on α for physiologically relevant parameters, provided that $\sigma \ll 1$. In the mathematical limit of $n_c \gg 2$, the first two terms of eq 14 can be neglected (because $\chi < 1$), leading to a simple solution $\chi = \min[1/(2\alpha), 1]$. Therefore the static mass fraction of monomers does not depend on σ and n_c in this limit. Our detailed numerical calculations further verify that this independence persists in the whole range of n_c, provided $\sigma \ll 1$ (Figure 2B).

From eq 13, we can also obtain the static number concentration of filaments as $P(\infty) = m_{tot}(1 - \chi)/[n_c + (n_c - 1)2\alpha\chi] \approx m_{tot}(1 - \chi)/(2n_c - 1)$, and the static average size of filaments as $M(\infty)/P(\infty) \approx 2n_c - 1$.

Thus, given the initial soluble protein concentration m_{tot} and critical nucleus size n_c, the static number and mass concentrations of filaments would mainly depend on the fiber elongation rate k_+ and the fragmentation rate k_- through a simple dimensionless form $\alpha = k_+ m_{tot}/[k_- n_c(n_c - 1)]$. This means the final static state has no apparent dependence on the nucleation process (except for the critical nucleus size n_c), which coincides with the general intuition about the thermodynamic systems.[35]

It must be noted that the static average size of filaments $(2n_c - 1)$ predicted above is inconsistent with general experimental results, in which much longer filaments and fibrils can be observed stable for days and even months. This discrepancy we believe is mainly caused by the unrealistic assumption on the fragmentation process in current model, where we suppose the fibrils can break at any position with equal probability and short fibrils can still break with significant probability. Yet both experiments and theoretical analysis[49] show that filaments can be more easily broken in the middle; and short filaments are much harder to break than long ones. Therefore further efforts are needed before the model can be applied to describe the end-stage of fibrillation.

In summary, the above outlined four successive phases (i.e., the initial lag phase, an exponential growth phase, a continuously breaking phase and the final static phase) provide a complete description of the formation of amyloid fiber. The former two stages have been studied extensively over the course of the past decade, whereas far less attention has been paid to the latter two. However, as we have noted here, the breaking phase may play a key role in the nosogenesis and aggravation of many amyloid-related diseases.

E. Classification of Amyloid Proteins. From the analysis of the exponential growth phase (eq 11), we can obtain a power-law dependence for the half-time and apparent growth rate of fiber formation on the initial protein concentration, i.e., $t_{1/2} \propto m_{tot}^{-\mu}$ and $k_{app} \propto m_{tot}^{\nu}$. When the secondary nucleation is much stronger than the primary nucleation ($\sigma \ll 1$), we have $\mu = \nu = 1/2$ (eq 11). Such weak dependencies have been observed in many amyloid proteins, including the yeast prion Sup35 NW region,[13] WW domain,[36] Csg B_{trunc},[37] Ure2 protein,[38] β2-microglobulin,[39] stefin B,[40] α-synucleins[31] and insulin,[41] as demonstrated in Figure 3.[51] If we compare the equations for the apparent growth rate and half-time of fiber formation, an inverse relationship $k_{app} \propto t_{1/2}^{-1}$ will be obtained (Figure 3E,F), which has been experimentally verified in many amyloid proteins regardless of the pH, temperature, and denaturant conditions.[42] Similar results could also be obtained for the lag-time.

However, we note that the scaling exponents μ and ν may show large deviations from $1/2$ for certain amyloid proteins.[43] As shown in Figure 4, $\mu = \nu = 1.5$ for γC-Crystallin,[44] and $\mu = \nu = 2.5$ for Apo C−II.[45] In fact, these values correspond to another limit case of our model at $\sigma \gg 1$, in which fragmentation is not as effective as the primary nucleation in providing seeds. Thus we have $\mu = \nu = n_c/2$, both depending on the size of critical nucleus.

As a result, it is natural to classify amyloid proteins into different groups based on the dimensionless parameter σ, which characterizes the effectiveness of primary nucleation against monomer-independent secondary nucleation (fragmentation) during the fibrillation process. More specifically, type I with $\sigma \ll 1$ corresponds to the case where protofilaments and mature filaments are both easily broken, so that fragmentation plays a key role in providing fiber seeds in the whole process. Most amyloid proteins in our study belong to this group. We have $\mu = \nu = 1/2$, which are independent of the critical nucleus size.
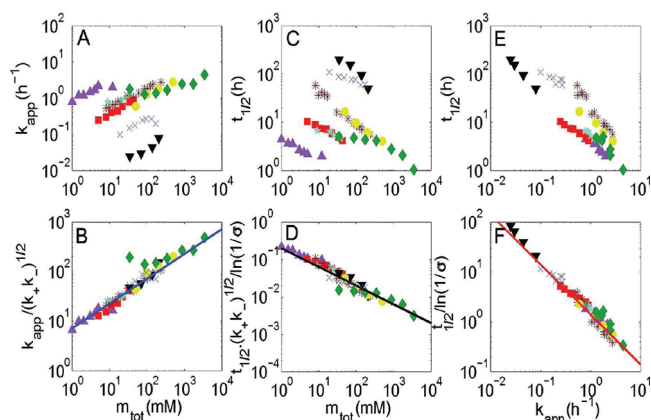
Figure 3. Scaling relationships between k_{app}, $t_{1/2}$, and m_{tot} for different kinds of amyloid proteins, whose scaling exponent does not depend on the critical nucleus size. Original data and corrected data by considering the factors predicted by the model are shown separately. (A) and (B) show the relationship between the apparent fiber growth rate k_{app} and initial protein concentration m_{tot} before and after correction, (C) and (D) show the relationship between half-time of fiber formation $t_{1/2}$ and initial protein concentration m_{tot} before and after correction, and (E) and (F) show the inverse relationship between the apparent fiber growth rate k_{app} and the half-time of fiber formation $t_{1/2}$ before and after correction. Here the data are obtained from eight different amyloid proteins, i.e., the yeast prion Sup35 NW region (purple triangles up), Csg B_{trunc} (red squares), Ure2 protein (cyan pentacles), β2-microglobulin (brown stars), stefin B (blue cross), α-synucleins (black triangles down), WW domain (yellow circles), and insulin (green diamonds). The blue fitting line is given by $y = 7.2x^{1/2}$, the black line is $y = 0.2x^{-1/2}$, and the red line is $y = \sqrt{2}x^{-1}$.
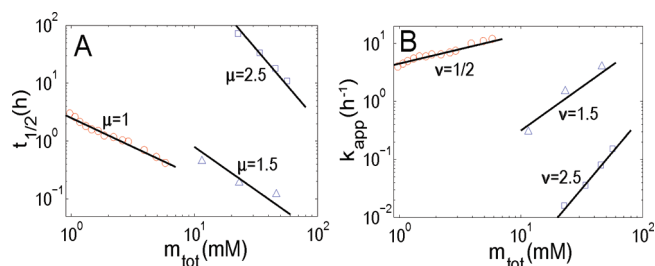

Figure 4. Scaling relationships between $t_{1/2}$, k_{app}, and m_{tot} are shown for three amyloid proteins, whose scaling exponent depends on the critical nucleus size. (A) shows the relationship between the half-time of fiber formation $t_{1/2}$ and initial protein concentration m_{tot}; (B) shows the relationship between the apparent fiber growth rate k_{app} and initial protein concentration m_{tot}. Here the data are collected for Aβ(M1−42) (red circles), γC-Crystallin (blue triangles), and Apo C-II (purple rectangles). The slope of each fitting curve is noted by the plot separately.

Type II with $\sigma \gg 1$ corresponds to the case where protofilaments and mature filaments are hard to break, and therefore primary nucleation dominates the process of fiber formation. Under this situation, we have $\mu = \nu = n_c/2$, which depends on the critical nucleus size, e.g., γC-Crystallin and Apo C-II.

Type III with $\sigma \sim 1$ corresponds to the case when primary nucleation dominates the initial process and fragmentation becomes more important in later stages as filaments get easier to break upon elongation. Thus it is an intermediate type between types I and II, with $\mu \approx n_c/2$ and $\nu \approx 1/2$ (e.g., Aβ(M1−42),
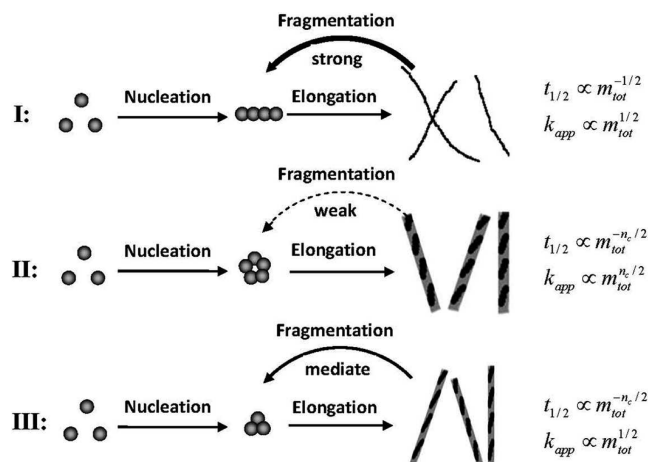

Figure 5. Illustration of the classification of amyloid proteins based on their different dominant nucleation mechanisms.

$\mu = 1$, $\nu = 1/2$ obtained from the numerical fitting of experimental data in ref 46).

In fact, types I and II of amyloid proteins follow the Oosawa theory[35] and the monomer-independent secondary nucleation models developed by Ferrone,[47] Miranker,[48] Dobson,[30] and others, whereas type III is new. A summary of above classifications is illustrated in Figure 5.

It should be noted that in nature high powers of the scaling exponents may not necessarily correspond to the systems that are dominated by primary nucleation. As shown in the case of sickle hemoglobin polymerization,[47,48] such high exponents can also originate from monomer-dependent secondary nucleation, which has not been considered in our model. How to quantitatively distinguish these two types of nucleation mechanisms is still an interesting open problem.

To provide further validation of our model, we take a look at the fibrillation processes of two types of amyloid proteins, Csg B_{trunc}[37] and Apo C-II.[45] It can be seen that, with fixed values of model parameters k_+, k_-, k_n, and n_c, the predicted kinetic curves provide an excellent agreement with the experimental data at low initial protein concentrations. At higher concentrations there is a small difference due to the modest influence of initial protein concentration on model parameters. Therefore we believe that the master equation of eq 3 correctly reflects the fundamental kinetic behaviors of amyloid fiber formation, in spite of the disparate governing nucleation mechanisms for different types of amyloid proteins.

■ SUMMARY AND DISCUSSION

On the basis of the asymptotic analysis of the master equation, we found that the kinetic process of amyloid fiber formation can be generally divided into four successive stages: lag phase, exponential growth phase, breaking phase and a final static phase. Each stage is characterized by its distinct time scale, the number and mass concentrations of filaments etc. We also emphasized the power-law dependence of the half-time and apparent growth rate of fiber formation on the initial protein concentration, which eventually leads to a tentative classification of amyloid proteins according to their different dominant nucleation mechanisms. We hope that detailed experiments can be conducted to validate the model, in particular the proposed classification of amyloid proteins and their corresponding fibrillation mechanisms in the future.
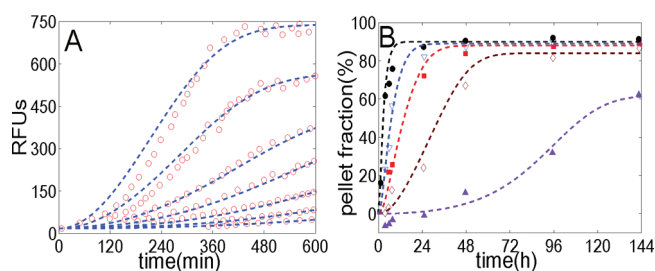
6615

dx.doi.org/10.1021/jp205702u | J. Phys. Chem. B 2012, 116, 6611−6617

Figure 6. Data analysis for the fibrillation of two different kinds of amyloid proteins: Csg B_{trunc} and apo C-II. (A) The red circles represent the measured fluorescence intensity of Csg B_{trunc} at concentrations of 43, 33, 22, 16, 11, 8, and 5 μM (from top to bottom), which was mixed with the amyloid-specific dye thioflavin T. We note that none of the pellet fractions reach 100% probably due to the way of normalization. Our fits (blue dashed lines) are calculated with eq 3 with parameters $k_+ = 1 \times 10^4\,M^{-1}\,s^{-1}$, $k_- = 2 \times 10^{-8}\,s^{-1}$, $k_n = 1.8 \times 10^{-4}\,M^{-1}\,s^{-1}$, and $n_c = 2$. (B) Symbols correspond to the experimental data monitored by a pelleting assay with apo C-II at protein concentrations of 0.15 mg/mL (purple triangles up), 0.3 mg/mL (brown diamonds), 0.45 mg/mL (red squares), 0.6 mg/mL (blue triangles down), and 0.9 mg/mL (black circles). Our fits (dashed lines) are obtained with parameters $k_+ = 5 \times 10^3\,M^{-1}\,s^{-1}$, $k_- = 1 \times 10^{-9}\,s^{-1}$, $k_n = 3 \times 10^8\,M^{-1}\,s^{-1}$ and critical nucleus $n_c = 5$.

It was shown that primary nucleation, elongation, and fragmentation can be interpreted as the fundamental steps in amyloid fiber formation. However, this does not necessarily mean that other processes lack significance. As we pointed out, protein structure unfolding and refolding (or structure conversion) usually exists as a prerequisite step in the formation of amyloid fibrils. The importance of branching, merging, and thickening may vary from case to case. How to take these effects into consideration in our model frame remains to be elucidated. In the supporting materials, we tried to give a possible way to add the processes of protein structure conversion into the model.

Another issue concerns the appropriate determination of model parameters, because in reality the kinetic process of amyloid fiber formation is sensitive to protein sequence, temperature, solvent pH, denaturant concentration, etc. The explicit correlation between fibrillation conditions and model parameters will largely determine the ability of current model in interpreting experimental data and making new predictions. From numerical studies on the kinetic polymerization curves of different amyloid proteins, we found that for secondary nucleation dominated proteins ($\sigma \ll 1$), $n_c = 2$ seems work well, probably due to the insensitivity of secondary nucleation on the critical nucleus size. For primary nucleation dominated proteins ($\sigma \gg 1$), a larger value of n_c (usually 5−6) is needed for the correct fitting (Figures 4 and 6B). In the methodology part, an empirical method for the determination of (k_+, k_-, k_n) based on the time evolutionary curves of amyloid fibrillation is introduced. However, a final solution to this problem requires a detailed microscopic theory, which should take into account the explicit or empirical chemical information regarding to the fibrillation conditions.

■ ASSOCIATED CONTENT

ⓢ Supporting Information. Model inclusion of protein structure conversion and comparison of the estimated model parameters with the experimental data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: zhng@umich.edu.

■ REFERENCES

(1) Merlini, G.; Bellotti, V. *New Engl. J. Med.* **2003**, *349* (6), 583–596.
(2) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
(3) Harrison, R. S.; Sharpe, P. C.; Singh, Y.; Fairlie, D. P. *Rev. Physiol. Biochem. Pharmacol.* **2007**, *159*, 1–77.
(4) Feng, B. Y.; Toyama, B. H.; Wille, H.; Colby, D. W.; Collins, S. R.; May, B. C.; Prusiner, S. B.; Weissman, J.; Shoichet, B. K. *Nat. Chem. Biol.* **2008**, *4* (3), 197–199.
(5) Roberts, B. E.; Shorter, J. *Nat. Struct. Mol. Biol.* **2008**, *15* (6), 544–546.
(6) Frieden, C. *Protein Sci.* **2007**, *16* (11), 2334–2344.
(7) Serio, T. R.; Cashikar, A. G.; Kowal, A. S.; Sawicki, G. J.; Moslehi, J. J.; Serpell, L.; Arnsdorf, M. F.; Lindquist, S. L. *Science* **2000**, *289* (5483), 1317–1321.
(8) van der Linden, E; Venema, P. *Curr. Opin. Colloid Interface Sci.* **2007**, *12*, 158–165.
(9) Powers, E. T.; Powers, D. L. *Biophys. J.* **2008**, *94* (2), 379–391.
(10) Lomakin, A.; Chung, D. S.; Benedek, G. B.; Kirschner, D. A.; Teplow, D. B. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (3), 1125–1129.
(11) Hills, R. D.; Brooks, C. L. *J. Mol. Biol.* **2007**, *368* (3), 894–901.
(12) Morris, A. M.; Watzky, M. A.; Agar, J. N.; Finke, R. G. F *Biochemistry* **2008**, *47* (8), 2413–2427.
(13) Collins, S. R.; Douglass, A.; Vale, R. D.; Weissman, J. S. *PLoS Biol.* **2004**, *2* (10), 1582–1590.
(14) Pallitto, M. M.; Murphy, R. M. *Biophys. J.* **2001**, *81* (3), 1805–1822.
(15) Hill, E. K.; Krebs, B.; Goodall, D. G.; Howlett, G. J.; Dunstan, D. E. *Biomacromolecules* **2006**, *7* (1), 10–13.
(16) Xue, W. F.; Hellewell, A. L.; Gosal, W. S.; Homans, S. W.; Hewitt, E. W.; Radford, S. E. *J. Biol. Chem* **2009**, *284* (49), 34272–34282.
(17) Andersen, C. B.; Yagi, H.; Manno, M.; Martorana, V.; Ban, T.; Christiansen, G.; Otzen, D. E.; Goto, Y.; Rischel, C. *Biophys. J.* **2009**, *96* (4), 1529–1536.
(18) van Gestel, J.; de Leeuw, S. W. *Biophys. J.* **2006**, *90* (9), 3134–3145.
(19) Manno, M.; Craparo, E. F.; Podesta, A.; Bulone, D.; Carrotta, R.; Martorana, V.; Tiana, G.; San Biagio, P. L. *J. Mol. Biol.* **2007**, *366* (1), 258–274.
(20) Sunde, M.; Blake, C. C. F. *Q. Rev. Biophys.* **1998**, *31* (1), 1–+.
(21) Lomakin, A.; Teplow, D. B.; Kirschner, D. A.; Benedek, G. B. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (15), 7942–7947.
(22) Arnaudov, L. N.; de Vries, R. *J. Chem. Phys.* **2007**, *126* (14), 145106.
(23) Wetzel, R. *Acc. Chem. Res.* **2006**, *39* (9), 671–679.
(24) Padrick, S. B.; Miranker, A. D. *Biochemistry* **2002**, *41* (14), 4694–4703.
(25) Zhang, J. N.; Muthukumar, M. *J. Chem. Phys.* **2009**, *130* (3), 035102-1–035102-17.
(26) Lee, C. C.; Nayak, A.; Sethuraman, A.; Belfort, G.; Mcrae, G. J. *Biophys. J.* **2007**, *92* (10), 3448–3458.
(27) Hall, D.; Edskes, H. *J. Mol. Biol.* **2004**, *336* (3), 775–786.

(28) Kunes, K. C.; Cox, D. L.; Singh, R. R. P. *Phys. Rev. E* **2005**, *72* (5), 051915-1–051915-8.

(29) Smith, J. F.; Knowles, T. P. J; Dobson, C. M.; MacPhee, C. E.; Welland, M. E. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (43), 15806–15811.

(30) Knowles, T. P. J; Waudby, C. A.; Devlin, G. L.; Cohen, S. I. A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. *Science* **2009**, *326* (5959), 1533–1537.

(31) Uversky, V. N.; Li, J.; Fink, A. L. *J. Biol. Chem.* **2001**, *276* (47), 44284–44296.

(32) Librizzi, F.; Rischel, C. *Protein Sci.* **2005**, *14* (12), 3129–3134.

(33) Kodaka, M. *Biophys. Chem.* **2004**, *107* (3), 243–253.

(34) Ono, K.; Condron, M. M.; Teplow, D. B. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (35), 14745–14750.

(35) Oosawa, F.; Asakura, S. *Thermodynamics of the Polymerization of Protein*; Academic Press: New York, 1975.

(36) Ferguson, N.; Berriman, J.; Petrovich, M.; Sharpe, T. D.; Finch, J. T.; Fersht, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (17), 9814–9819.

(37) Hammer, N. D.; Schmidt, J. C.; Chapman, M. R. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (30), 12494–12499.

(38) Zhu, L.; Zhang, X. J.; Wang, L. Y.; Zhou, J. M.; Perrett, S. *J. Mol. Biol.* **2003**, *328* (1), 235–254.

(39) Xue, W. F.; Homans, S. W.; Radford, S. E. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (26), 8926–8931.

(40) Skerget, K.; Vilfan, A.; Pompe-Novak, M.; Turk, V.; Waltho, J. P.; Turk, D.; Zerovnik, E. *Bioinformatics* **2009**, *74* (2), 425–436.

(41) Nielsen, L.; Khurana, R.; Coats, A.; Frokjaer, S.; Brange, J.; Vyas, S.; Uversky, V. N.; Fink, A. L *Biochemistry* **2001**, *40* (20), 6036–6046.

(42) Fandrich, M. *J. Mol. Biol.* **2007**, *365* (5), 1266–1270.

(43) Wright, C. F.; Teichmann, S. A.; Clarke, J.; Dobson, C. M. *Nature* **2005**, *438* (7069), 878–881.

(44) Wang, Y. T.; Petty, S.; Trojanowski, A.; Knee, K.; Goulet, D.; Mukerji, I.; King, J. *Invest. Ophth. Vis. Sci.* **2010**, *51* (2), 672–678.

(45) Binger, K. J.; Pham, C. L. L.; Wilson, L. M.; Bailey, M. F.; Lawrence, L. J.; Schuck, P.; Howlett, G. J. *J. Mol. Biol.* **2008**, *376* (4), 1116–1129.

(46) Hellstrand, E.; Boland, B.; Walsh, D. M.; Linse, S. *ACS Chem. Neurosci.* **2010**, *1* (1), 13–18.

(47) Ferrone, F. A.; Hofrichter, J.; Eaton, W. A. *J. Mol. Biol.* **1985**, *183* (4), 611–631.

(48) Ruschak, A. M.; Miranker, A. D. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (30), 12341–12346.

(49) Hill, T. L. *Biophys. J.* **1983**, *44*, 285–288.

(50) Equation 5 can be directly derived from the definitions of κ, σ, and α in eq 9, whereas eq 6 can be obtained from the asymptotic solutions of $t_{1/2}$, k_{app}, and χ in eq 11.

(51) We can observe that the data plots for some proteins in Figure 3 are not so linear, such as insulin and α-synuclein. The main reason is that the fibrillation processes under high protein concentration and low protein concentration can be quite different from each other, which has not been considered in current model.

(52) A problem with eq 2 is the production of fibrils smaller than the critical nucleus during fragmentation. Such fibrils cannot grow by monomer addition because, according to the model, only size $i \geq n_c$ can grow. They will be frozen in these states if $n_c > 2$. In a correct treatment, the second and third terms accounting for fiber fragmentation should be modified as $-k_-(j - 2n_c + 1)f(t,j)$ and $+2k_-\Sigma_{i=j+n_c} f(t,i)$. However, here for easy comparison, we still take the original model as in Knowles et al. paper.

6617

dx.doi.org/10.1021/jp205702u |*J. Phys. Chem. B* 2012, 116, 6611–6617