Stochastic calculus of protein filament formation under spatial confinement

Thomas C T Michaels\textsuperscript{1,2}, Alexander J Dear\textsuperscript{1} and Tuomas P J Knowles\textsuperscript{1,3}

\textsuperscript{1} Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
\textsuperscript{2} Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, United States of America
\textsuperscript{3} Cavendish Laboratory, Department of Physics, University of Cambridge, J J Thomson Avenue, Cambridge, CB3 0HE, United Kingdom

E-mail: tctm3@cam.ac.uk

Keywords: nucleation, stochastic differential equation, lag-time, growth rate, amyloid, cell, microdroplets

Abstract
The growth of filamentous aggregates from precursor proteins is a process of central importance to both normal and aberrant biology, for instance as the driver of devastating human disorders such as Alzheimer’s and Parkinson’s diseases. The conventional theoretical framework for describing this class of phenomena in bulk is based upon the mean-field limit of the law of mass action, which implicitly assumes deterministic dynamics. However, protein filament formation processes under spatial confinement, such as in microdroplets or in the cellular environment, show intrinsic variability due to the molecular noise associated with small-volume effects. To account for this effect, in this paper we introduce a stochastic differential equation approach for investigating protein filament formation processes under spatial confinement. Using this framework, we study the statistical properties of stochastic aggregation curves, as well as the distribution of reaction lag-times. Moreover, we establish the gradual breakdown of the correlation between lag-time and normalized growth rate under spatial confinement. Our results establish the key role of spatial confinement in determining the onset of stochasticity in protein filament formation and offer a formalism for studying protein aggregation kinetics in small volumes in terms of the kinetic parameters describing the aggregation dynamics in bulk.

1. Introduction

Protein filament formation is an example of homomolecular self-assembly of central importance to normal biology, disease, and current nanotechnology. For instance, biofilaments of the proteins actin and tubulin play pivotal roles in the regulation of cell growth, cell movement and cell division [1–11]. Many bacteria also exploit protein filaments during biofilm formation [12]. However, deregulated protein filament aggregation can be implicated in a number of human disorders [13–16]. Key examples of such pathologies include Alzheimer’s disease, Parkinson’s disease, sickle-cell anemia and type-II diabetes, which are associated with the formation in the human brain or other organs of filamentous protein assemblies commonly known as amyloid fibrils [13–24]. Finally, due to their unique physicochemical properties, self-assembled filamentous structures can be used as biomaterials for many applications in nanotechnology [25–31].

Theoretical investigations of protein filament formation [1, 2, 32–37] are commonly based upon the mean-field limit of the law of chemical kinetics [38], which neglects potential effects from intrinsic noise. While in vitro protein filament formation reactions at the μl-scale (bulk) are generally well-explained by these deterministic mechanistic models (figure 1(a)) [1, 2, 32–37], protein aggregation kinetics in very small volumes (≈1 μl) is known to yield non-reproducible reaction curves (figure 1(c)) [39–45], indicating that for such systems the stochastic nature of the underlying assembly process plays a fundamental role in determining the time course of aggregation. Such a scenario would be realized, for instance, inside living cells (figure 1(d)) [44], which typically involve compartmentalization in far smaller reaction volumes (≈1 fl–1 μl) than those commonly employed in conventional experiments in bulk (≈1 μl). Furthermore, droplet microfluidics [45] now permit the
study of protein aggregation in reaction volumes comparable to intracellular scales ($\approx 1$ nl–1 pl), further contributing to the need to understand the role of volume confinement in protein filament assembly (figure 1(e)).

Here, we introduce a kinetic framework based on stochastic calculus [46, 47] that allows us to investigate the stochastic behavior of protein filament formation processes in small volumes. Starting from a purely deterministic description of filamentous growth kinetics [32–37], we derive a stochastic differential equation governing the early-time evolution of protein filament formation under spatial confinement. We exploit this formalism to describe the statistical properties of individual small-volume protein aggregation reactions and derive explicit expressions for the probability distribution of lag-times. Moreover, we establish the increasing importance of rare primary nucleation events in controlling the characteristic properties of aggregation in the stochastic limit.

### 2. Deterministic aggregation kinetics

The dynamics of protein filament formation in bulk is modeled in terms of a set of deterministic kinetic equations, the mean-field master equation [38], that describes the time evolution of the concentrations of aggregates of different sizes in terms of the elementary mechanisms of aggregation. The fundamental microscopic events that drive protein filament formation (figure 1(b)) generally involve an initial primary nucleation step, where the smallest fibrillar units (nuclei) are formed spontaneously from monomers, coupled to growth through filament elongation at the ends of filaments. In many cases, including the aggregation of disease-related peptides and proteins such as Alzheimer’s Amyloid-$\beta$ peptide, the aggregation process is accelerated by secondary nucleation processes, whereby new fibrils can be formed through the interaction between monomers and existing fibrils [32, 35, 36, 48]. The most commonly experimentally accessible observables include the monomer concentration $m(t)$ as well as the number and mass concentrations of fibrils, denoted with $P(t)$ and $M(t)$. These quantities are defined in terms of the concentrations $f(t,j)$ of aggregates of size $j$ as $P(t) = \sum_j f(t, j)$ and $M(t) = \sum_j j f(t, j)$. The general dynamic equations that describe the time evolution of $P(t)$ and $M(t)$ for a system evolving through primary and secondary nucleation pathways are thus obtained by summation of the
mean-field master equation over aggregation number $j$, yielding [32–37]:

$$\frac{dP(t)}{dt} = k_1 m(t)^{n_1} + k_2 m(t)^{n_2} M(t),$$  
(2.1)

$$\frac{dM(t)}{dt} = 2k_1 m(t) P(t),$$  
(2.2)

where $k_1$, $k_2$ and $k_\lambda$ are the rate constants for primary nucleation, secondary nucleation and filament elongation and $n_1$ and $n_2$ are the reaction orders for primary and secondary nucleation with respect to the monomer concentration. Note that equation (2.2) tacitly assumes aggregate growth to be the dominant term contributing to changes in aggregate mass concentration. This assumption, however, is always satisfied in filamentous aggregating systems, because the formation of long aggregates necessarily requires nucleation processes to be slow compared to aggregate growth.

### 2.1. Early stages of aggregation

Kinetic equations such as (2.1) and (2.2) describe a sigmoidal-type kinetics characterized by an initial lag-phase, followed by rapid growth and a final phase of approach to the plateau (figure 1(a)) [49]. The system dynamics during the initial stages of aggregation (lag-phase and rapid growth) are captured very accurately by assuming that the monomer concentration is not significantly depleted, i.e. $m(t) \approx m(0)$, where $m(0)$ is the initial concentration of monomers. Under these circumstances, the deterministic aggregation kinetics are captured by a simpler set of linearized kinetic equations, obtained from equations (2.1) and (2.2) in the limit of constant monomer concentration, $m(t) \approx m(0)$ [32–37]:

$$\frac{dP(t)}{dt} = \alpha + \beta M(t),$$  
(2.3)

$$\frac{dM(t)}{dt} = \mu P(t),$$  
(2.4)

where we have introduced the notation $\alpha = k_1 m(0)^{n_1}$, $\beta = k_2 m(0)^{n_2}$ and $\mu = 2k_1 m(0)$. In this regime, both the filament number and mass concentrations then increase exponentially with time, $P(t) \approx \alpha/(2\kappa)e^{\mu t}$ and $M(t) \approx \alpha/(2\beta)e^{\kappa t}$ after a short initial adjustment period, where $\kappa = \sqrt{\mu \beta} = \sqrt{2k_1 k_2 m(0)^{n_2+1}}$ is the fibril multiplication rate. Hence, $P(t)$ and $M(t)$ are proportional to each other, $P(t) = (\beta/\kappa) M(t)$; equivalently, the average fibril length is constant in this regime, at $L = \kappa/\beta$. Through this proportionality relationship, equations (2.3) and (2.4) can be simplified further to give a single and linear differential equation that describes the early-time evolution of filament mass concentration $M(t)$ [41, 50]:

$$\frac{dM(t)}{dt} = \kappa M(t) + \xi,$$  
(2.5)

where $\xi = \kappa k_1 m(0)^{n_1-n_2}/(2k_1)$. Equation (2.5) has a straightforward physical interpretation: $\xi$ is an input term describing the rate at which new filaments are introduced in the system through primary nucleation, while the parameter $\kappa$ describes an effective rate of filament multiplication through secondary nucleation. We have effectively mapped our system onto one without explicit elongation, such that all filaments are of length $L$. New filaments of this length are added by both the input term, and the multiplication term.

### 3. Stochastic aggregation kinetics

The kinetic approach for describing the growth kinetics of protein filaments discussed in the previous section is based on the assumption that the growth rate $\kappa$ and the input parameter $\xi$ entering equation (2.5) are perfectly known quantities. In the following, we generalize equation (2.5) by accounting for the effects of intrinsic noise associated with small-volume effects.

#### 3.1. Stochastic differential equation model for aggregation in small volumes

To see how system volume $V$ enters explicitly equation (2.5) and determines the onset of stochasticity, it is useful to note that the parameter $\kappa$ has units of inverse time and, thus, the first term on the right-hand side of equation (2.5) is independent of system volume. By contrast, the input term $\xi$, which describes the rate of increase of aggregate mass due to primary nucleation, has units of concentration and inverse time and hence carries an intrinsic dependence on $V$. In particular, the number of primary nucleation events per unit time is expressed as $\lambda = k_1 m(0)^{n_1} N_A V$, where $N_A$ is Avogadro’s number, while the quantity $\nu = \xi/\lambda$ represents the increase in aggregate mass concentration associated with a single primary nucleation event. Thus, reducing the system volume $V$ leads to few and rare primary nucleation events. Consequently, aggregation curves become highly unpredictable because variations in the actual number of primary nucleation events that occur, and the
For the left limit of the paper, although the left limit will always be tacitly assumed.

As a technical note, it is important to clarify that equation (46, 47) must be understood as $dM(t) = \kappa M(t^-) dt + \nu dN(t)$, where $f(t^-)$ stands for the left limit of $f$, i.e. $f(t^-) = \lim_{t \downarrow s} f(t)$. This clarification is important as it highlights the fact that the fluxes depend on the value of $M(t)$ immediately before the jump, thus ensuring causality. For notational simplicity, however, the subscript $t^-$—will be dropped throughout the paper, although the left limit will always be tacitly assumed.
3.2. Stochastic aggregation curves

The explicit solution to the stochastic differential equation (3.1) can be found from direct integration as

\[ M(t) = e^{\nu t} \int_0^t K(u) e^{-\nu s} dN(s). \tag{3.2} \]

This expression can be re-written using the definition of stochastic integrals for jump processes \cite{46,47}, such that the solution to equation (3.1) can be obtained explicitly as:

\[ M(t) = \nu \sum_{n=1}^{N(t)} e^{\nu(t-T_n)}, \tag{3.3} \]

where the stochastic process \( N(t) \) has jumps at random times \( T_n \).

Equation (3.3) describes in closed-form the stochastic time evolution of a reaction \( M(t) \) having monodisperse initial conditions and Poisson white noise input. The term \( e^{\nu(t-T_n)} \) entering the sum in equation (3.3) describes the deterministic propagation of aggregate mass following a stochastic nucleus formation event that occurred at time \( T_n \). Hence, the measured stochastic polymerization curves explicitly depend on the current number \( N(t) \) of nuclei that happened to have formed at time \( t \) and the times \( T_n \) at which these nuclei were formed. The variability of the measured \( M(t) \) profiles thus originates because \( N(t) \) is a stochastic process and the times \( T_n \) are random. Figure 2(a) shows 10 sample aggregation curves generated using equation (3.3). Clearly, all realizations of \( M(t) \) differ from each other because both \( N(t) \) and \( T_n \) are random. Equation (3.3) therefore describes irreproducible kinetics.

3.3. Statistics of stochastic aggregation curves

Using equation (3.3), we can now derive closed-form expressions for the average properties of the stochastic kinetic curves. The expected value of \( M(t) \) is given by
Note that in the deterministic limit, \( H \) can be written as a sum of independent \( \lambda \) Poisson processes, due to the \( \nu \) factor. Hence, for the variance of the \( \nu \) term of the variance is \( \lambda - \nu \) decreasing volume, increases exponentially in time and the extent of fluctuations in the measured aggregation curves increases with decreasing volume, \( \text{Var}[M(t)] \approx V^{-2} \). The coefficient of variation, \( c_v = \sqrt{\text{Var}[M(t)]}/\mathbb{E}[M(t)] \), is asymptotically constant on long times.

### 3.4. Lag-time distribution

A key feature of linear protein polymerization reactions is the observation of an initial lag-phase during which no aggregation is detected. For reproducible reaction curves, quantification of the delay in the polymerization is typically given in terms of a lag-time \( \tau(x) \), which, for a reaction starting with a concentration \( M(0) = x \) of seed aggregates, is defined as the time at which \( M(t) \) reaches an arbitrarily chosen concentration threshold \( M_{th} \).

\[
M(\tau(x)) = M_{th}.
\] (3.11)
Hence, the deterministic value of the lag-time can be determined from (2.5) and reads
\[
\tau(x) = \frac{1}{\kappa} \log \left( \frac{M_{th} + \xi/\kappa}{x + \xi/\kappa} \right).
\]  
(3.12)

For reaction kinetics exhibiting stochastic behavior, however, the duration of the lag-phase can no longer be defined as in equation (3.11), because the individual kinetic curves are not reproducible. Instead, the lag-time \( \tau \) becomes a random variable, whose variability reflects the fluctuations in the underlying polymerization reaction [46, 51]. Under these circumstances, finding \( \tau \) translates into a first hitting time problem. We define \( \tau(x) \) as the time at which a sample path of the stochastic process \( M(t) \) first hits the barrier \( M_{th} \)
\[
\tau(x) := \inf_{t \geq 0} \{ M(t) \geq M_{th} | M(0) = x \leq M_{th} \}. 
\]  
(3.13)

The underlying mechanism of barrier hitting is sketched in figure 2(b). Of particular interest are the cumulative probability distribution function of the lag-time, defined as
\[
Q(t, x) := \mathbb{P}[\tau(x) \leq t] = \mathbb{P}[\max_{0 \leq s \leq t} M(s) \leq M_{th} | M(0) = x]
\]  
(3.14)

for all \( t \geq 0 \), and the associated probability density function
\[
T(t, x) = \frac{dQ(t, x)}{dt}.
\]  
(3.15)

In fact, the theory of exit times for Markov processes [46, 51] allows us to re-formulate the first passage problem in terms of a partial differential equation with appropriate initial and boundary conditions. In particular, the cumulative probability distribution function of lag-times for a reaction starting at \( M(0) = x \), \( Q(t, x) \), satisfies the backward Kolmogorov equation:
\[
\frac{\partial Q(t, x)}{\partial t} = LQ(t, x)
\]  
(3.16)

subject to the boundary condition
\[
Q(t, x) = 1, \quad x \geq M_{th}
\]  
(3.17)

and the initial condition
\[
Q(0, x) = 0, \quad x < M_{th}.
\]  
(3.18)

Here, \( L \) is the backward Kolmogorov operator of the SDE 3.1, which is given by\(^5\)
\[
Lf(x) = \kappa x \frac{\partial f(x)}{\partial x} + \lambda[f(x + \nu) - f(x)].
\]  
(3.22)

Noticing that jump step sizes due to primary nucleation are very small compared to \( M_{th} \), i.e. \( \nu/M_{th} \ll 1 \), we can re-write \( L \) most conveniently as follows:
\[
Gf(x) \approx (\kappa x/\nu + \lambda)[f(x + \nu) - f(x)],
\]  
(3.23)

where, we have replaced the first derivative of \( f(x) \) with the corresponding finite difference. Equation (3.16) can now be solved in Laplace space, where the Laplace transform \( \hat{Q}(s, x) = \int_0^\infty e^{-st} Q(t, x) dt \) of \( Q(t, x) \) satisfies the following recurrence equation
\[
(\kappa x/\nu + \lambda + s)\hat{Q}(s, x) = (\kappa x/\nu + \lambda)\hat{Q}(s, x + \nu),
\]  
(3.24)

with boundary condition \( \hat{Q}(s, M_{th}) = 1/s \). The solution is
\[
\hat{Q}(s, 0) = \frac{1}{s} \sum_{j=0}^{\theta-1} \frac{(\lambda + j\kappa)}{(\lambda + j\kappa + s)},
\]  
(3.25)

where \( \theta = M_{th}/\nu \). Inverse Laplace-transforming equation (3.25) yields the following closed form solution for the lag-time cumulative probability distribution function

\(^5\) Consider a jump-diffusion process described by the following SDE
\[
dX(t) = \mu(X) dt + \sigma(X) dW(t) + \nu dN(t),
\]  
(3.19)

where \( W(t) \) is a Wiener process and \( N(t) \) is a Poisson jump process with intensity \( \lambda \) and constant jump size \( \nu \). Then, the probability distribution \( p(x, t) \) that \( X(t) = x \) satisfies the forward Kolmogorov equation
\[
\frac{\partial p(x, t)}{\partial t} = \frac{\partial}{\partial x} [\mu(x)p(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2}[\sigma(x)^2 p(x, t)] + \lambda [p(x-\nu, t) - p(x, t)].
\]  
(3.20)

The adjoint of the forward Kolmogorov equation is known as backward Kolmogorov equation and reads:
\[
\frac{\partial f(x, t)}{\partial t} = \mu(x) \frac{\partial f(x, t)}{\partial x} + \frac{\sigma(x)^2}{2} \frac{\partial^2 f(x, t)}{\partial x^2} + \lambda [f(x+\nu, t) - f(x, t)].
\]  
(3.21)
According to equation (3.15), the lag-time probability density function $T(t, 0)$ is finally obtained by differentiating equation (3.26), yielding [42, 50]:

$$T(t, 0) = \frac{k\Gamma(\lambda/\kappa + \theta)}{\Gamma(\theta)\Gamma(\lambda/\kappa)} e^{-\lambda t} (1 - e^{-\kappa t})^{\theta - 1}. \tag{3.27}$$

Thus, by explicit calculation of the random aggregation trajectories we verified the results derived previously in [42, 50].

3.5. Characteristics of lag-time distribution and average lag-time

Equation (3.27) describes in closed form the probability distribution of lag-times for a reaction with no initial aggregates (figure 2(d)). In the long time limit, the probability density function for lag-times decays exponentially

$$T(t, 0) \overset{t\to\infty}{\to} \frac{k\Gamma(\lambda/\kappa + \theta)}{\Gamma(\theta)\Gamma(\lambda/\kappa)} e^{-\lambda t}, \tag{3.28}$$

whereby the characteristic decay time is $1/\lambda$. In the early-time limit, $T(t, 0)$ and all its derivatives decay to zero, i.e. $T(t, 0)$ has an essential singularity at $t = 0$. The function $T(t, 0)$ attains its maximum at

$$\tau^* = \frac{1}{\kappa} \log \left[ 1 + \frac{(\theta - 1)\kappa}{\lambda} \right]. \tag{3.29}$$

which can be interpreted as the most probable time at which the threshold concentration $M_{th}$ is reached. Since $\theta \gg 1$, this is effectively equal to the deterministic lag-time. The expected value for the lag-time, $\langle\tau\rangle$, is obtained by integrating equation (3.27), yielding

$$\langle\tau\rangle = \int_0^\infty T(t, 0) t \, dt = \sum_{j=0}^{\theta-1} \frac{1}{\lambda + j\kappa}. \tag{3.30}$$

The variance of the lag-time distribution is computed similarly and reads

$$\text{Var}(\tau) = \langle\tau^2\rangle - \langle\tau\rangle^2 = \sum_{j=0}^{\theta-1} \frac{1}{(\lambda + j\kappa)^2}. \tag{3.31}$$

Figure 2(c) illustrates a plot of $\langle\tau\rangle$ as a function of inverse system volume. In the limit of large volumes or high nucleation rate, corresponding to $\lambda \gg \kappa$, $\langle\tau\rangle$ tends to the deterministic bulk lag-time, or the most-probable lag-time

$$\langle\tau\rangle \to \int_0^\theta \frac{dx}{\lambda + \kappa x} = \frac{1}{\kappa} \log \left( \frac{M_{th} \kappa}{\xi} + 1 \right) = \tau_{\text{bulk}}. \tag{3.32}$$

In the opposite limit of small volumes or slow nucleation, corresponding to $\lambda \ll \kappa$, the sum in equation (3.30) is dominated by the $j = 0$ term. Consequently, $\langle\tau\rangle$ becomes

$$\langle\tau\rangle \to \frac{1}{\lambda} = \frac{c_m}{V}, \tag{3.33}$$

where $c_m^{-1} = k_m m(0)^{\kappa} N_0$. In the limit of small volumes, the rate of primary nucleation thus becomes so slow that the aggregation reaction is limited by the waiting time for the first few primary nucleation events to occur. In this regime, $\langle\tau\rangle$ displays a marked volume dependence. In particular, in a plot of $\langle\tau\rangle$ against $1/V$, equation (3.30) approaches the bulk limit $\tau_{\text{bulk}}$ approximately as a straight line of gradient $c_m$. Since $c_m$ depends explicitly on the rate constant for primary nucleation, $k_m$, a measurement of average lag-times with inverse system volume under spatial confinement provides a useful strategy for investigating the early stages of protein filament formation. These stages are otherwise highly challenging to probe using traditional bulk experiments, as aggregate concentration readouts in bulk tend to be dominated by secondary nucleation rather than primary nucleation. The transition between the bulk regime and equation (3.33) occurs at a characteristic volume $V_c = \kappa c_m$. Estimates of $V_c$ are typically in the picoliter range, which is many orders of magnitude smaller than reaction volumes in bulk. Microfluidic techniques have emerged as a powerful method for probing such volumes in a high throughput manner. Distinct small-volume aggregation reactions may be carried out (and measured) in parallel in large numbers of microdroplets, generating sufficient statistics to determine average lag-times as a function of volume.
3.6. Growth rate

Another observable of key experimental interest is the normalized average growth rate at time \( t \), which in the protein aggregation literature is commonly defined as:

\[
r(t) = \frac{1}{\langle M(t) \rangle} \frac{dM(t)}{dt}.
\]

Note that other definitions of the growth rate are equally possible, including, for instance, \( r(t) = 1/t \log(\langle M(t) \rangle) \).

The growth rate (3.34) can be evaluated by considering the averaged version of equation (3.1)

\[
\left\langle \frac{dM(t)}{dt} \right\rangle = \kappa \langle M(t) \rangle + \nu \left\langle \frac{dN(t)}{dt} \right\rangle.
\]

To evaluate the term \( \left\langle \frac{dN(t)}{dt} \right\rangle \), we first note that, since \( N(t) \) is a sum of equally-sized step functions with steps at times \( T_n \), its derivative must be a uniformly-weighted sum of delta functions at the same times:

\[
\frac{dN(t)}{dt} = \sum_{n=1}^{\infty} \delta(t - T_n).
\]

Hence, since the jumps are independent, and their times \( T_n \) are Gamma-distributed (see table 1), we find:

\[
\left\langle \frac{dN(t)}{dt} \right\rangle = \sum_{n=1}^{\infty} x^n e^{-x} \tau^{n-1} \frac{\tau^n}{(n-1)!} = \lambda.
\]

Since \( \nu \lambda = \xi \), it follows:

\[
r(t) = \frac{1}{\langle M(t) \rangle} \left\langle \frac{dM(t)}{dt} \right\rangle = \kappa + \frac{\xi}{\langle M(t) \rangle}.
\]

Because \( \langle M(t) \rangle \approx e^{rt} \), for times \( t \gg \kappa^{-1} \) the expression (3.38) can be simplified to:

\[
r(t) = \kappa + \frac{\xi}{\langle M(t) \rangle} \approx \kappa.
\]

Equation (3.39) states that the normalized growth rate equals \( \kappa \) irrespective of system volume, \( V \). This is because, once an appreciable fraction of monomer has aggregated, the processes that control further growth, i.e. elongation and secondary nucleation, are deterministic in nature.

3.7. Correlation between normalized growth rate and lag-time

In the deterministic limit, there is a strong linear correlation between inverse lag-time and normalized growth rate [33]. This correlation has been verified experimentally for different systems in bulk [52, 53] and emerges because both the inverse lag-time and the normalized growth rate scale with \( \kappa \); moreover, the proportionality constant between inverse lag-time and normalized growth rate is in the form of a logarithmic correction that depends only weakly on different experimental conditions. Our results show that this correlation is lost gradually as system volume \( V \) is decreased and the system enters the stochastic regime (figure 3). This is because the normalized growth rate still equals \( \kappa \) in the stochastic limit, but the lag-time \( \tau \) tends to \( 1/\lambda \). This effect reflects that the lag-time no longer represents solely the time for taken for \( M(t) \) to multiply up to the threshold value, but also the time for the first primary nucleation events to occur. Thus, reducing the system volume...
decouples the effect of the primary nucleation process from that of the other growth processes, permitting its direct measurement.

4. Conclusions

In this paper, we have studied stochastic effects in protein aggregation phenomena in small volumes. Using the framework of stochastic calculus with jump processes, we have obtained expressions that describe the stochastic time course of protein aggregation reactions and the probability distribution of lag-times. We have further demonstrated the gradual loss of correlation between protein aggregation rate and lag-times as the system size shrinks. Our results highlight the crucial role played by the system volume in determining the inherent variability characterizing protein aggregation phenomena under confinement and provide a strategy for linking bulk parameters characterizing large-volume kinetics with behavior in small volumes.

Acknowledgments

We acknowledge support from Peterhouse, Cambridge (TCTM), the Swiss National Science foundation (TCTM), the Schiff Foundation (AID), the Wellcome Trust (TPJK), the Cambridge Centre for Misfolding Diseases (TPJK), the BBSRC (TPJK), the Frances and Augustus Newman foundation (TPJK). The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) through the ERC grant PhysProt (agreement no. 337969).

Appendix. Calculation of $\mathbb{E}[M(t)^2]$

The second moment of the polymerization curve distribution is computed as:

$$\mathbb{E}[M(t)^2] = \nu^2 e^{2\kappa t} \sum_{n=0}^{\infty} \frac{e^{-\lambda n}}{n!} \mathbb{E}\left[\sum_{i=1}^{n} e^{-\kappa (T_i + T_f)}\right]$$

The expectation value of the double sum in equation (A.1) can be re-written as:

$$\mathbb{E}\left[\sum_{i=1}^{n} \sum_{j=1}^{n} e^{-\kappa (T_i + T_f)}\right] = \sum_{k=1}^{n} \mathbb{E}[e^{-2\kappa T_k}] + \sum_{i=k}^{n} \mathbb{E}[e^{-\kappa T_i}] \mathbb{E}[e^{-\kappa T_k}]$$

$$= \sum_{k=1}^{n} r_k^2 + r_k^2 \sum_{i=1}^{n} r_i^k - r^{2k} \mathbf{(A.2)}$$

where $r_k = \lambda/(\lambda + 2\kappa)$ and $r = \lambda/(\lambda + \kappa)$. Noting that:

$$\sum_{i=1}^{n} r_i^k = \frac{r}{1 - r}(1 - r^n); \quad \frac{r}{1 - r} = \frac{\lambda}{\kappa}; \quad \frac{r_2}{1 - r_2} = \frac{\lambda}{2\kappa},$$

we find:

$$\mathbb{E}\left[\sum_{i=1}^{n} \sum_{j=1}^{n} e^{-\kappa (T_i + T_f)}\right] = \frac{\lambda}{2\kappa} + \frac{\lambda^2}{\kappa^2} - \frac{\lambda^2}{\kappa^2 + 2\kappa \lambda} - \frac{\lambda r_2}{2\kappa} - 2 \frac{\lambda^2}{\kappa^2} r^n$$

$$+ \left(\frac{\lambda^2}{\kappa^2} + \frac{\lambda^2}{\kappa^2 + 2\kappa \lambda}\right) r^{2n}. \mathbf{(A.4)}$$

Inserting equation (A.4) into (A.1), we find:

$$\mathbb{E}[M(t)^2] = \nu^2 e^{2\kappa t} \sum_{n=0}^{\infty} \frac{e^{-\lambda n}}{n!} \mathbb{E}\left[\sum_{i=1}^{n} \sum_{j=1}^{n} e^{-\kappa (T_i + T_f)}\right]$$

$$= \left(\frac{\xi^2}{2\kappa^2} + \frac{\xi^2}{\kappa^2 + 2\kappa \lambda}\right) e^{2\kappa t} - \frac{\xi^2}{2\kappa^2} e^{\lambda t + 2\kappa t}$$

$$- 2 \frac{\xi^2}{\kappa^2} e^{\lambda + 2\kappa t} + \left(\frac{\xi^2}{\kappa^2} + \frac{\xi^2}{\kappa^2 + 2\kappa \lambda}\right) e^{\lambda + \kappa t + 2\kappa t}. \mathbf{(A.5)}$$

An alternative to the direct calculation of $\mathbb{E}[M(t)^2]$ would be to transform equation (3.1) into a stochastic differential equation for $M(t)^2$ using Ito’s lemma and then to average this equation in order to obtain a differential equation for the expected value $\mathbb{E}(M(t)^2)$.
References

[1] Oosawa F and Asakura S 1975 Thermodynamics of the Polymerization of Proteins (Waltham, MA: Academic)
[2] Oosawa F and Kasai M 1962 J. Mol. Biol. 4 110
[3] Tobacman J S and Korn E D 1983 J. Biol. Chem. 258 3207
[4] Frieden C and Goddette D W 1983 Biochemistry 22 5836
[5] Wegner A and Engel J 1975 Biophys. Chem. 3 215
[6] Fowler D M, Koulou A V, Alory-Jost C, Marks M S, Balch W E and Kelly J W 2006 PLAS Biol. A 6
[7] Fowler D M, Koulou A V, Balch W E and Kelly J W 2007 Trends Biochem. Sci. 32 217
[8] Kelly J W and Balch W 2003 J. Cell Biol. 161 461
[9] Mostaert A S, Higgins M J, Fukuma T, Rindi F and Jarvis S P 2006 J. Biol. Phys. 32 393
[10] Watt B et al 2009 J. Biol. Chem. 284 35543
[11] Mitchinson T and Kirschner M 1984 Dynamic instability of microtubule growth Nature 312 237–42
[12] Maji S K et al 2009 Functional amyloids as natural storage of peptide hormones in pituitary secretory granules Science 325 328–32
[13] Chiti F and Dobson C M 2006 Annu. Rev. Biochem. 75 333
[14] Chiti F and Dobson C M 2017 Annu. Rev. Biochem. 86 27
[15] Dobson C M 2017 The amyloid phenomenon and its links with human disease Cold. Spring. Harb. Perspect. Biol. 9 a023648
[16] Knowles T P J, Vendruscolo M and Dobson C M 2014 The amyloid state and its association with protein misfolding diseases Nat. Rev. Mol. Cell Biol. 15 384–96
[17] Dobson C M 2003 Nature 426 884
[18] Prusiner S B 1991 Science 252 1513
[19] Come J H, Fraser P E and Lansbury P T 1993 Proc. Natl Acad. Sci. USA 90 5859
[20] Aguzzi A and O’Connor T 2010 Protein aggregation diseases: pathogenicity and therapeutic perspectives Nat. Rev. Drug Discov. 9 237
[21] Li C, Adamcik J and Mezzenga R 2012 Adv. Drug Deliv. Rev. 64 14
[22] Hardy J and Selkoe D J 2002 Science 297 353
[23] Selkoe D J 2003 Nature 426 900
[24] Lansbury P T and Lashuel H A 2006 Nature 443 774
[25] Gazit E 2007 Chem. Soc. Rev. 36 1263
[26] Knowles T P J and Buehler M J 2011 Nat. Nanotechnol. 6 469
[27] Knowles T P J, Oppenheim T, Buell A K, Chirgadze D Y and Welland M E 2010 Nat. Nanotechnol. 5 204
[28] Li C, Adamcik J and Mezzenga R 2012 Nat. Nanotechnol. 7 421
[29] Li C, Bolisetty S and Mezzenga R 2013 Adv. Mater. 25 3694
[30] Li C, Alam M M, Bolisetty S, Adamcik J and Mezzenga R 2011 Chem. Commun. 47 2913
[31] Bolisetty S, Adamcik J, Heier J and Mezzenga R 2012 Adv. Funct. Mater. 22 3424
[32] Bishop M F and Ferrone F A 1984 Biophys. J. 46 631
[33] Knowles T P J, Waudby C A, Devlin G L, Cohen S I A, Aguzzi A, Vendruscolo M, Terentjev E M, Welland M E and Dobson C M 2009 Science 326 1533
[34] Cohen S I A, Vendruscolo M, Welland M E, Dobson C M, Terentjev E M and Knowles T P J 2011 J. Chem. Phys. 135 065105
[35] Cohen S I A et al 2013 Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism Proc. Natl Acad. Sci. USA 110 9758–63
[36] Meisel G et al 2016 Molecular mechanisms of protein aggregation from global fitting of kinetic models Nat. Protocols 11 252–72
[37] Michaels T C T, Cohen S I A, Vendruscolo M, Dobson C M and Knowles T P J 2016 Phys. Rev. Lett. 116 038101
[38] Michaels T C T and Knowles T P J 2014 Am. J. Phys. 82 476
[39] Ferrone F A, Hofrichter J and Eaton W A 1980 Biophys. J. 32 361
[40] Ferrone F A, Hofrichter J and Eaton W A 1985 J. Mol. Biol. 183 611–31
[41] Hofrichter J 1986 J. Mol. Biol. 189 553
[42] Michaels T C T, Dear A J, Kirkegaard J B, Saar K L, Weitz D A and Knowles T P J 2016 Phys. Rev. Lett. 116 258103
[43] Szavits-Nossan J, Eden K, Morris R J, MacPherson C E, Evans M R and Allen R J 2014 Phys. Rev. Lett. 113 098101
[44] Colby D W, Cassidy J P, Lin G C, Ingram V M and Wittrup K D 2006 Nat. Chem. Biol. 2 319
[45] Knowles T P J, White D A, Abate A R, Agresti J J, Cohen S I A, Sperling R A, De Genst E J, Dobson C M and Weitz D A 2011 Proc. Natl Acad. Sci. USA 108 14746
[46] Cox D R and Miller H D 1965 The Theory of Stochastic Processes (New York: Wiley)
[47] Grigoriou M 2002 Stochastic Calculus (Boston, MA: Birkhäuser)
[48] Saric A, Buell A K, Meisel G, Michaels T C T, Dobson C M, Linse S, Knowles T P J and Frenkel D 2016 Nature 12 874
[49] Arosoio F, Knowles T P J and Linse S 2015 Phys. Chem. Chem. Phys. 17 7606
[50] Szabo A 1988 J. Mol. Biol. 199 539
[51] Tuckwell H and Wan F Y M 1984 J. Appl. Probab. 21 695
[52] Zhu L, Zhang X-J, Wang L-Y, Zhou J-M and Perrett S 2003 J. Mol. Biol. 328 235
[53] Fändrich M 2007 J. Mol. Biol. 365 1266