THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Modeling Protein Aggregation Kinetics: The Method of Second Stochasticization

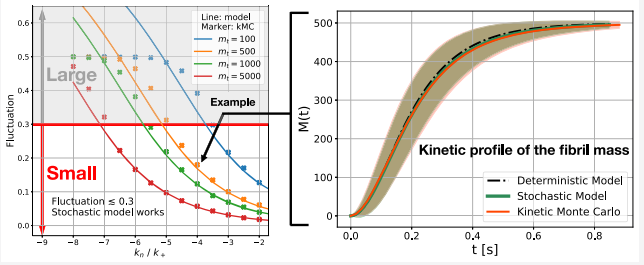
Published as part of The Journal of Physical Chemistry virtual special issue "Yoshitaka Tanimura Festschrift".

Jia-Liang Shen,[§] Min-Yeh Tsai,[*,§] Nicholas P. Schafer, and Peter G. Wolynes[*]

Cite This: *J. Phys. Chem. B* 2021, 125, 1118–1133

Read Online

ABSTRACT: The nucleation of protein aggregates and their growth are important in determining the structure of the cell's membraneless organelles as well as the pathogenesis of many diseases. The large number of molecular types of such aggregates along with the intrinsically stochastic nature of aggregation challenges our theoretical and computational abilities. Kinetic Monte Carlo simulation using the Gillespie algorithm is a powerful tool for modeling stochastic kinetics, but it is computationally demanding when a large number of diverse species is involved. To explore the mechanisms and statistics of aggregation more efficiently, we introduce a new approach to model stochastic aggregation kinetics which introduces noise into already statistically averaged equations obtained using mathematical moment closure schemes. Stochastic moment equations summarize succinctly the dynamics of the large diversity of species with different molecularity involved in aggregation but still take into account the stochastic fluctuations that accompany not only primary and secondary nucleation but also aggregate elongation, dissociation, and fragmentation. This method of "second stochasticization" works well where the fluctuations are modest in magnitude as is often encountered *in vivo* where the number of protein copies in some computations can be in the hundreds to thousands. Simulations using second stochasticization reveal a scaling law that correlates the size of the fluctuations in aggregate size and number with the total number of monomers. This scaling law is confirmed using experimental data. We believe second stochasticization schemes will prove valuable for bridging the gap between *in vivo* cell biology and detailed modeling. (The code is released on https://github.com/MYTLab/stoch-agg.)



I. INTRODUCTION

The aggregation of proteins into large molecular assemblies is crucial for cell biology: some large aggregates are necessary for proper function, while others cause disease. For example, aggregates of G-actin filaments are a major component of the dynamic cytoskeleton of cells, while amyloid-β (Aβ) and τ-protein aggregates are associated with Alzheimer's disease.[1,2] In the past few decades, considerable theoretical effort has been devoted to the physical chemistry of protein aggregation. A major difficulty in understanding aggregation is that a wide range of species is involved, all having different numbers of component particles as well as myriad possible shapes. In 1962, Oosawa and his co-workers proposed a simple model to investigate the kinetics of the growth of linear and helical aggregates of proteins.[3] They employed the principal moment method to characterize the distribution of species in order to derive a closed-form solution of the deterministic rate equations. The results of their analysis agreed well with experiment. Later, they modified their model by adding the process of primary nucleation, where monomers must form nuclei before polymerization can occur.[4] In 1985, the importance of a secondary nucleation channel was realized by Eaton and co-workers and was used to describe the lag time observed in the polymerization of sickle hemoglobin[5] using a perturbation method.[6] More recently, Knowles et al. have proposed holistic mechanistic models that include other secondary processes, such as filament breakage,[7] which can promote secondary nucleation. All of these kinetic rate models, being deterministic descriptions of average behavior, work quite well for macroscopic experiments *in vitro* and provide useful insights into aggregation mechanisms.[8,9]

We must acknowledge however that cells and their compartments are microscopic so stochastic effects are prominent. Stochastic models of aggregation in cells must be able to deal with the same large variety of processes found *in vitro*: primary/secondary nucleation, elongation, and dissocia-

Received: November 16, 2020
Revised: January 4, 2021
Published: January 21, 2021

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118–1133

tion, as well as secondary processes, such as fragmentation.[10] These processes can be described through a set of equations for all the species involved.[8] It is best to compare the solutions of these equations with experimental kinetics by monitoring suitable averages over the size distributions. These averages connect to the relevant experimental probes. Key averages include the total number of monomers found in aggregates which is proportional to the total mass in aggregates, as well as the second principal moment of the distribution of species sizes. Both of these quantities are commonly used to monitor the progress of aggregation. Note that the second principal moment of the distributions refers to the total mass in aggregates ($M(t) = \sum_{j=n_c}^{\infty} jf(t, j)$, where $f(t, j)$ is the number of length-j aggregates). By globally fitting $M(t)$ to the experimental data *in vitro*, one can obtain fitted values for the kinetic parameters (e.g., critical nucleus size, nucleation/elongation rate constant) and thus gain mechanistic insight into the relative importance of the different microscopic processes.[11]

Single cell experiments performed *in vivo* or microfluidics experiments using small volumes highlight the inherent variability of the aggregation process.[12] In these situations, the time course of aggregation displays significant random fluctuation due to the small number of protein molecules that can aggregate. These number fluctuations are quite important in the cell. The deterministic moment closures, however, do not account for such fluctuations. In a single cell, the number of monomers is sufficiently small such that the intrinsic fluctuations become physiologically significant. Ferrone et al. were the first to discuss the stochastic fluctuations of the lag time for aggregation in small volumes.[5] Since then, many studies on the fluctuations of the kinetics of aggregation have appeared.[12,13] A dominant theoretical approach[13] for modeling follows the full probability distribution for each species over time. By solving the master equation for the probability distribution of aggregates, one obtains the early time distribution of the fibril mass and can also calculate the fluctuations of the lag time. Knowles et al.[14] used a different scheme to model the stochasticity of the early phases of the kinetics. Their approach effectively employed a stochastic differential equation (SDE) with jump noise to quantify the discrete changes in the fibril mass. They assumed a constant monomer concentration in order to obtain the closed form early time solution. Fluctuations, however, also play an important role in the later stages of aggregation once monomers have been depleted. In this work, we develop a stochastic approach that allows the efficient calculation of the statistics of the kinetic profile for all times. This approach resembles the chemical Langevin equation approach.[15−17] These equations are nonlinear because they employ a derivative matching moment closure technique[18] to numerically calculate the moments of the polymeric species distributions for all times. By introducing Gaussian white noises associated with the relevant underlying elementary chemical reaction events into the rate equations for the concentrations of the individual filament sizes $f(t, j)$, the number of length-j aggregates, we are able to derive an equation for the fluctuations of the principal moments of the filament size distribution. This procedure leads to a manageable problem that can now be described using only two coupled stochastic differential equations. We describe this strategy as "second stochasticization" in analogy to "second quantization" in quantum field theory, where we quantize again a set of (supposedly already averaged) wave fields. One of the nonlinear stochastic equations describes the time course of the first principal moment, which is the total number of aggregates $P(t) = \sum_{j=n_c}^{\infty} f(t, j)$. The other equation describes the dynamics of the second principal moment of the distribution, which measures the total number of monomers found in the aggregate form $M(t) = \sum_{j=n_c}^{\infty} jf(t, j)$ (or the fibril mass). In the model, a total of five mutually statistically independent Gaussian white noises need to be introduced. These noise terms reflect the stoichiometries of the various nucleation and growth processes. We will focus on the case where the typical size of the fibrils that are formed is much larger than the size of the critical nucleus, reflecting the inclusion of both primary and secondary nucleation. The stochastic differential equations that result from this procedure are nonlinear, and thus the mathematical expressions for the lowest population moments are not fully closed, i.e., the mean and variance of $P(t)$ and $M(t)$ couple to still higher moments of the size distributions. The derivative matching moment closure technique offers an organized but approximate way to close the moment hierarchy. There are also other ways to approximate the statistics of these nonlinear systems of stochastic differential equations, such as first passage methods,[19] perturbation methods, or numerical sampling. These other approaches are not our current focus.

The advantage of the second stochasticization strategy is that it provides a much more efficient route to getting the full-time stochastic kinetics in comparison with kinetic Monte Carlo (kMC) simulation of all the species that are involved, especially when the total number of monomers is very large. The computational complexity for running conventional kMC simulations scales at least linearly with the system's complexity[20] ($\sim O(N)$ with N indicating the number of reaction channels). In this paper, we do however benchmark the approximated solutions with the results obtained from the kMC simulations. Second stochasticization results agree quite well with the kMC simulations in the moderate fluctuation regime.

The rest of the paper is organized as follows. In section II, we briefly review the most common deterministic aggregation model and the corresponding principal moment-based mathematical expressions. In section III, we describe the concentration fluctuations by stochastic differential equations (SDEs); i.e., we derive the chemical Langevin method. This then leads to a set of coupled ordinary differential equations (ODEs) for the principal moments of the size distribution using Ito calculus. Owing to the nonlinear nature of nucleation processes and fragmentation, the resulting moment equations are not closed. We therefore use a moment closure technique to derive a final set of second stochasticized equations for the dynamics. In section IV, we compare the resulting stochastic model with the results from the kMC simulations. This comparison yields a scaling law that correlates the total number of monomers m_t with the strength of the stochastic fluctuations. This scaling relation is then confirmed using experimental data. Second stochasticization provides an efficient and accurate description of the aggregation kinetic profile and its fluctuations in time.

II. DETERMINISTIC AGGREGATION KINETICS

In this section, we review the simplest kinetic model of aggregation. The model is essentially a coarse-grained model of

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

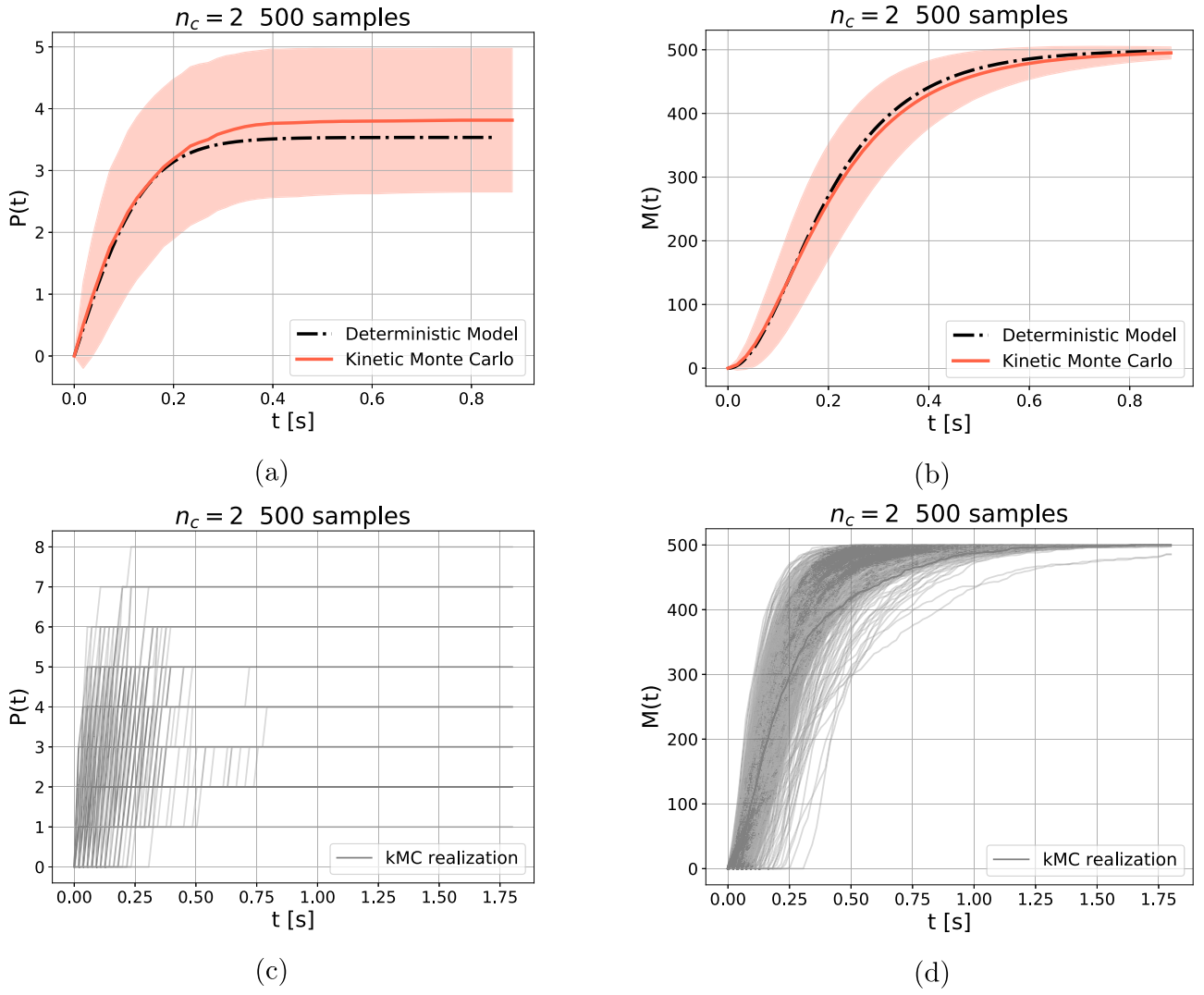


Figure 1. Kinetic profiles of $P(t)$ and $M(t)$ obtained from both the kinetic Monte Carlo (kMC) simulations and from the deterministic model. The upper panels (a, b) compare the deterministic model with the kMC simulations. (a) $P(t)$. (b) $M(t)$. The dashed black lines represent the results from the deterministic model. For kMC simulation, the mean and the standard deviation (SD) of all of the samples are shown using the red solid lines and pink shadows, respectively. Lower panels (c, d) display the corresponding kinetic MC realizations of parts a and b, respectively. Each gray line represents individual kinetic trajectory from independent simulations. For 500 kinetic MC samples, the kinetic parameters used are $k_n = 10^{-4}$ s^{-1} and $k_+ = 1$ s^{-1}; Other parameters include $m_t = 500$, $P(0) = 0$, $n_c = 2$, $k_- = 10^{-6}$ s^{-1}, $k_f = 10^{-8}$ s^{-1}, and $k_2 = 0$.

a well-mixed system that shows the full-time evolution of the number of aggregates of all sizes. The time course of aggregation depends on a competition between many distinct molecular processes. To aggregate, free monomers must form an unstable nucleus through a primary nucleation reaction before any further reactions can take place. Elongation reactions add monomers onto an existing nucleus or shorter aggregate, while dissociation reactions release monomers from an aggregate into the solution environment. Fragmentation and secondary nucleation also generate more nuclei and thus accelerate aggregation. Fragmentation does this by breaking an aggregate into two separate pieces, thereby increasing the number of aggregates with open ends at which further monomers can bind. If, however, the fragment that breaks off is smaller than the nucleus size, it melts away, so in that case fragmentation does not necessarily produce more aggregated species. Secondary nucleation describes a catalyzed heterogeneous nucleation process, where the surface of existing long aggregates catalyzes additional nucleation events. Note that in the classical deterministic model, while all these reactions occur at random, they are averaged over in the kinetic equations so that stochasticity is not explicitly accounted for.[8]

II.A. Dynamics of the Species Distribution. The average aggregation kinetics can be described through a set of rate equations for the number of different species, each made up of a different number of monomeric units. The quantity $f(t, j)$ denotes the number of aggregates made up j units that are found at time t. When all the processes are included, the time course of $f(t, j)$ can be written as follows[8] (please note that $j \geq n_c$):

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

$$\frac{\partial f(t,j)}{\partial t} = 2m(t)k_+ f(j-1) - 2m(t)k_+ f(j) + 2k_- f(j+1) - 2k_- f(j) + 2k_f \sum_{i=j+1}^{\infty} f(i) - k_f(j-1)f(j) + k_n m(t)^{n_c}\delta_{j,n_c} + k_2 m(t)^{n_2} \sum_{i=n_c}^{\infty} if(t,i)\delta_{j,n_2}. \quad (1)$$

In this equation, $m(t)$ is the number of free monomers, [k_n, k_+, k_-, k_f, k_2] are rate coefficients for primary nucleation, elongation, dissociation, fragmentation, and secondary nucleation, respectively, and n_c is the parameter that describes the apparent size of the critical nucleus needed for primary nucleation. The two terms in the first line describe the elongation process. The first term reflects the fact that when a monomer is added onto a length-$(j-1)$ aggregate ($A_{j-1} + A_1 \rightarrow A_j$), one forms a new length-j aggregate, while the second term refers to the addition of a monomer onto an existing length-j aggregate ($A_j + A_1 \rightarrow A_{j+1}$), which thus leads to a decrease in the number of length-j aggregates. In the present paper, we do not distinguish growth at either end so the factor of 2 arises because each aggregate has two ends available for elongation or dissociation into monomers. The second line accounts for dissociation reactions, through which an aggregate loses a monomer. Fragmentation is described by the third line: the first term corresponds with the increase of the number of length-j aggregates due to the breakage of a long length-i ($i > j$) aggregate into a length-j aggregate ($A_i \rightarrow A_j + A_{i-j}$), while the second term represents the fragmentation of a length-j aggregate into a length-k aggregate ($j > k$) ($A_j \rightarrow A_{j-k} + A_k$), which can occur at $j-1$ possible sites. Primary nucleation is encoded in the fourth line ($n_c A_1 \rightarrow A_{n_c}$), which shows up when $j = n_c$, an equivalent of the Kronecker delta (δ_{j,n_c}). In this work, we will assume that aggregates with a length of $1 < j < n_c$ are unstable so that they either move on to aggregate or break into monomers rapidly. In other words, we assume there is no appreciable concentration of intermediates with size less than n_c.

II.B. Moment Equations. Several approximation schemes have been proposed to solve the deterministic eq 1 analytically, and among these methods, the method of principal moments is the most commonly used. The zeroth principal moment, $P(t) = \sum_{j=n_c}^{\infty} f(t,j)$, is the total number of aggregates in the system, and the first principal moment, $M(t) = \sum_{j=n_c}^{\infty} jf(t,j)$, counts the total number of monomers that have aggregated or the total mass of aggregates (the fibril mass). These two principal moments of $f(t,j)$ are closely related to the observables that can be measured in bulk experiments. The moment equations are

$$\frac{dP(t)}{dt} = k_f[M(t) - (2n_c - 1)P(t)] + k_n m(t)^{n_c} + k_2 m(t)^{n_2} M(t)$$

$$\frac{dM(t)}{dt} = 2m(t)k_+ P(t) - 2k_- P(t) - k_f n_c(n_c - 1)P(t) + n_c k_n m(t)^{n_c} + n_2 k_2 m(t)^{n_2} M(t). \quad (2)$$

The first equation dP/dt describes the time evolution of the total number of aggregates, while the second equation describes the time evolution of the total mass of aggregates. Higher principal moments of $f(t,j)$ have also been investigated in the literature,[9] but the first two moments, $P(t)$ and $M(t)$, are generally considered to be the most important quantities because of their direct connection to experimental observables. Analytical solutions of eq 1 and 2 have been discussed.[21−23] These solutions have been successfully used to characterize several bulk experimental results[24,25] and have provided mechanistic insights for several different aggregation scenarios, for example, an extended model for the size-dependent aggregation and fragmentation rates.[26]

Parts a and b of Figure 1 display the kinetic profiles of $P(t)$ and $M(t)$, respectively; we compare the results obtained from both the deterministic moment equations and from kinetic Monte Carlo (kMC) simulations. The deterministic model shows a single kinetic curve that should be compared to the statistical mean value of $P(t)$. Although this kinetic curve agrees well with the mean value obtained from the kMC simulations, it gives us no idea concerning the size of the stochastic fluctuation effects. The kMC realizations of $P(t)$ and $M(t)$ make clear the stochastic nature of the aggregation trajectories for finite size samples. These traces are shown in parts c and d of Figure 1, respectively. The fluctuations are not significant when the number of initial monomers m_t is large, but they become non-negligible when m_t is small. The case where the system has initially only a small number of free-to-nucleate monomers (e.g., $m_t = 10^2$−10^4) is of particular interest because this size range represents a typical number of copies of proteins that would be found *in vivo* in a human cell or compartment (≈ 500 fL).[12] These numbers were estimated by directly scaling the known protein concentrations (∼1−10^2 nM) to such a compartment. For cells, then, stochastic fluctuations are not negligible. Fluctuations also affect the true averaged kinetic curves significantly. The calculations shown do not use the "constant monomer assumption", which is generally not valid for cells. To understand the dynamics of cells, stochasticity must explicitly be incorporated into the model. We do this using the Langevin description of the underlying kinetic laws with fluctuations.

III. STOCHASTICITY IN THE KINETICS OF THE AGGREGATION MODEL

The chemical Langevin equation approach[15,16,27] can be used to describe the intrinsic fluctuations of aggregation processes. The chemical Langevin equation describes the fluctuations of the number of each species, X_i, by adding *noise* terms into the ordinary deterministic rate equations, thus forming a stochastic differential equation (SDE). The general form of the chemical Langevin equation is

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M} \nu_{ji} a_j(\mathbf{X}(t)) + \sum_{j=1}^{M} \nu_{ji} a_j^{1/2}(\mathbf{X}(t))\Gamma_j(t) \quad (3)$$

The index j runs over all chemical reactions in which the species X_i participates. The change in the number of i particles X_i, due to the jth reaction is represented by the stoichiometry factor ν_{ji}. $a_j(\mathbf{X}(t))$ is the propensity function for the jth reaction. The first term in eq 3 is the usual deterministic rate equation for X_i. The last term represents the noise that must be added to those rate equations in order to reflect the individuality of reaction events. $\Gamma_j(t)$ is a set of statistically

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

independent Gaussian white noises. Owing to the assumed instantaneous nature of the reactions, the noise terms have no correlation in time, $\langle \Gamma_j(t)\Gamma_{j'}(t')\rangle = \delta_{jj'}\delta(t-t')$.

III.A. The Stochastic Rate Equations for Aggregation. To illustrate the scheme, here, we write out explicitly eq 3 for the example of a specific elongation reaction step, $A_{18} + A_1 \rightarrow A_{19}$. If we want to write down the rate equation for the number of length-19 aggregates, $\mathrm{d}f(t, 19)/\mathrm{d}t$, we need to know the stoichiometry factors and propensities first. In a single molecular elongation step, the population of A_{19} changes by +1, thus the stoichiometry factor is 1. As for the propensity, we can write down the rate $2k_+m(t)f(t, 18)$, where the factor "2" arises because there are two sites available for elongation to take place. The noise term that is associated with the elongation reaction, $A_{18} + A_1 \rightarrow A_{19}$, in the equation $\mathrm{d}f(t, 19)/\mathrm{d}t$ is therefore

$$+1 \times \sqrt{2k_+m(t)f(t, 18)}\,\xi(A_{18} + A_1 \rightarrow A_{19})$$

. To avoid confusion, we will label the noise terms ξ with their corresponding associated reaction.

We can now proceed to derive the stochastic rate equation for the full aggregation system by adding corresponding noise terms to each of the rate equations for the aggregation reactions. Again, the noises are labeled by the type of the reaction, with A_j representing a length-j aggregate. The complete set of stochastic rate equations for $f(t, j)$ therefore reads

$$\begin{aligned}
\frac{\partial f(t,j)}{\partial t} &= 2m(t)k_+f(j-1) + \sqrt{2m(t)k_+f(j-1)}\,\xi(A_{j-1} + A_1 \rightarrow A_j) - 2m(t)k_+f(j) - \sqrt{2m(t)k_+f(j)}\,\xi(A_j + A_1 \rightarrow A_{j+1}) \\
&+ 2k_-f(j+1) + \sqrt{2k_-f(j+1)}\,\xi(A_{j+1} \rightarrow A_1 + A_j) - 2k_-f(j) + \sqrt{2k_-f(j)}\,\xi(A_j \rightarrow A_1 + A_{j-1}) + k_nm(t)^{n_c}\delta_{j,n_c} \\
&+ \sqrt{k_nm(t)^{n_c}}\,\delta_{j,n_c}\xi(n_cA_1 \rightarrow A_{n_c}) + k_2m(t)^{n_2}\sum_{i=n_c}^{\infty} if(t,i)\delta_{j,n_2} + \sqrt{k_2m(t)^{n_2}\sum_{i=n_c}^{\infty} if(t,i)}\,\delta_{j,n_2}\xi(n_2A_1 \rightarrow A_{n_2}) + 2k_f\sum_{i=j+1}^{\infty} f(i) \\
&+ \sum_{i=j+1, i\neq 2j}^{\infty}\sqrt{2k_ff(i)}\,\xi(A_i \rightarrow A_j + A_{i-j}) + 2\sqrt{k_ff(2j)}\,\xi(A_{2j} \rightarrow 2A_j) - (j-1)k_ff(j) \\
&+ \begin{cases} -\sqrt{2k_ff(j)}\sum_{i=1}^{(j-1)/2}\xi(A_j \rightarrow A_{j-i} + A_i) & j \text{ odd} \\ -\sqrt{2k_ff(j)}\sum_{i=1}^{(j-2)/2}\xi(A_j \rightarrow A_{j-i} + A_i) - \sqrt{k_ff(j)}\,\xi(A_j \rightarrow 2A_{j/2}) & j \text{ even} \end{cases}
\end{aligned} \tag{4}$$

For simplicity, we do not write down the time dependences of $f(j)$ and ξ explicitly. One can verify stoichiometric conservation is preserved in eq 4 by comparing, for instance, the terms in the rate equations $\partial f(2j)/\partial t$ and $\partial f(j)/\partial t$ that are associated with the $A_{2j} \rightarrow 2A_j$ fragmentation reaction,

$$\begin{aligned}
\left.\frac{\partial f(t,2j)}{\partial t}\right|_{A_{2j}\rightarrow 2A_j} &= -k_-f(2j) - \sqrt{k_-f(2j)}\,\xi(A_{2j} \rightarrow 2A_j) \\
\left.\frac{\partial f(t,j)}{\partial t}\right|_{A_{2j}\rightarrow 2A_j} &= 2k_-f(2j) + 2\sqrt{k_-f(2j)}\,\xi(A_{2j} \rightarrow 2A_j).
\end{aligned} \tag{5}$$

So it is clear that we get

$$-\left.\frac{\partial f(t,2j)}{\partial t}\right|_{A_{2j}\rightarrow 2A_j} = \frac{1}{2}\left.\frac{\partial f(t,j)}{\partial t}\right|_{A_{2j}\rightarrow 2A_j} \tag{6}$$

Equation 6 shows that the total number of monomers is conserved and that the change in the number of each species agrees with the stoichiometry of all reactions in which they participate. Notice that in the above equation, the noises that are introduced on both sides of the equations are identical and therefore are highly correlated with each other, since they refer to the same Gaussian noise.

III.B. The Stochastic Moment Equations. Following the same procedure as was used for the deterministic kinetic equations, we now derive the stochastic moment equations for averages over the number distributions. This procedure yields two coupled now stochastic differential equations, one for $P(t)$ and the other for $M(t)$ (see Appendix). The equations are

$$\begin{aligned}
\frac{\mathrm{d}P(t)}{\mathrm{d}t} &= k_nm(t)^{n_c} + \sqrt{k_nm(t)^{n_c}}\,\xi_{n_c} + k_2m(t)^{n_2}M(t) \\
&+ \sqrt{k_2m(t)^{n_2}M(t)}\,\xi_{n_2} \\
&+ k_f[M(t) - (2n_c - 1)P(t)] \\
&+ \sqrt{k_f[M(t) - (2n_c - 1)P(t)]}\,\xi_{fP} \\
\frac{\mathrm{d}M(t)}{\mathrm{d}t} &= n_ck_nm(t)^{n_c} + n_c\sqrt{k_nm(t)^{n_c}}\,\xi_{n_c} + n_2k_2m(t)^{n_2}M(t) \\
&+ n_2\sqrt{k_2m(t)^{n_2}M(t)}\,\xi_{n_2} + 2[m(t)k_+ - k_-]P(t) \\
&+ \sqrt{2[m(t)k_+ + k_-]P(t)}\,\xi_{\pm} - k_fn_c(n_c - 1)P(t) \\
&- \sqrt{k_f\frac{1}{3}n_c(n_c - 1)(2n_c - 1)P(t)}\,\xi_{fM}.
\end{aligned} \tag{7}$$

with the general noise term obeying $\langle \xi_M\xi_N\rangle = \delta_{MN}$, for $M, N \in \{n_c, n_2, \pm, fP, fM\}$. For example, $\xi_\pm$ refers to the sum of the noise in the elongation process, ξ_+, and the noise in the dissociation process, ξ_-. We can combine these two noise terms in the $M(t)$ equation into a single compound noise term because they do not appear in the $P(t)$ equation, and therefore this procedure leads to no complicated correlations between the noises for $P(t)$ and $M(t)$ (see the Appendix). The fragmentation noises shown in the two moment equations are labeled differently. As discussed in the Appendix, there is *almost* no correlation between the two types of fragmentation

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

noises, ξ_{fp} and ξ_{fM}, assuming that n_c is small compared to the average size of the aggregates, which is a condition that is almost always satisfied in protein aggregation systems. Typical fragmentation reactions change either the total number of aggregates when the reaction occurs in the middle region of an aggregate, or else change the fibril mass primarily (when the reaction takes place near the end of an aggregate), which results in releasing $n < n_c$ monomers back into the solution. Only when a fragmentation reaction changes both $P(t)$ and $M(t)$ will there be a correlation between ξ_{fP} and ξ_{fM}. In fact, this kind of fragmentation can only occur when a length l aggregate breaks, where $n_c < l < 2n_c$, resulting in losing one from $P(t)$ and losing l from $M(t)$. Note that this independence of the noises would be exact for the case of $n_c = 2$.

In this report, we will focus only on finding the mean and standard deviations (SD) of $P(t)$ and $M(t)$, and we will use these quantities to measure the stochastic fluctuations of the kinetic profiles observed in the kMC simulations as shown in Figure 1c,d. A key point, however, is that due to the nonlinearity of the stochastic moment equations, the moment equations do not strictly close. That is, the exact equations for a specific order of moment, $\langle X(t)^n\rangle$, also involve higher moments, such as $\langle X(t)^{n+1}\rangle$, and this kind of higher-moment-dependence continues to arbitrary order. To resolve this problem, one approach would be to generate samples from the equation, and then obtain the observables from the samples. Another approach is to truncate the higher-moment-dependence, stopping at a certain order say c, and express all the higher moments (those higher than c) in terms of the lower moments ($\leq c$). In the next section, we use such a moment closure method to obtain the approximated mean and the standard deviations of $P(t)$ and $M(t)$.

III.C. Moment Closure Method. The idea of the moment closure method is to express the higher moments in terms of products of powers of some lower moments. To investigate the time evolution of the mean and standard deviations of $P(t)$ and $M(t)$, or equivalently, $m(t)$ [$m(t) + M(t) = m_t$], we apply the moment closure method to the stochastic moment equation, eq 7, up to the second order (up to $\langle P(t)^\alpha m(t)^\beta\rangle$, where $\alpha + \beta = 2$). We then exploit the derivative matching moment closure technique[18] to obtain a set of ordinary differential equations (ODEs) for the retained moments, which can then be solved numerically. Here we show explicitly the set of ODEs that come from applying the moment closure method,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle P(t)\rangle &= k_n\langle m(t)^{n_c}\rangle + k_f\langle m_t - m(t) - (2n_c - 1)P(t)\rangle \\ &\quad + k_2\langle m(t)^{n_c}(m_t - m(t))\rangle\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle m(t)\rangle &= -n_c k_n\langle m(t)^{n_c}\rangle - 2k_+\langle m(t)P(t)\rangle + 2k_-\langle P(t)\rangle \\ &\quad + k_f n_c(n_c - 1)\langle P(t)\rangle - n_2 k_2\langle m(t)^{n_2}(m_t - m(t))\rangle\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle P(t)^2\rangle &= 2k_n\langle P(t)m(t)^{n_c}\rangle + k_n\langle m(t)^{n_c}\rangle \\ &\quad + 2k_f\langle P(t)m(t) - (2n_c - 1)P(t)^2\rangle \\ &\quad + k_f\langle m_t - m(t) - (2n_c - 1)P(t)\rangle \\ &\quad + 2k_2\langle P(t)m(t)^{n_2}(m_t - m(t))\rangle \\ &\quad + k_2\langle m(t)(m_t - m(t))\rangle\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle m(t)^2\rangle &= -2n_c k_n\langle m(t)^{n_c+1}\rangle + {n_c}^2 k_n\langle m(t)^{n_c}\rangle \\ &\quad - 4k_+\langle P(t)m(t)^2\rangle + 2k_+\langle P(t)m(t)\rangle + 4k_-\langle P(t)m(t)\rangle \\ &\quad + 2k_-\langle P(t)\rangle + 2k_f n_c(n_c - 1)\langle P(t)m(t)\rangle \\ &\quad + k_f\frac{1}{3}n_c(n_c - 1)(2n_c - 1)\langle P(t)\rangle \\ &\quad - 2n_2 k_2\langle m(t)^{n_2+1}(m_t - m(t))\rangle \\ &\quad + {n_2}^2 k_2\langle m^{n_2}(m_t - m(t))\rangle\end{aligned}$$

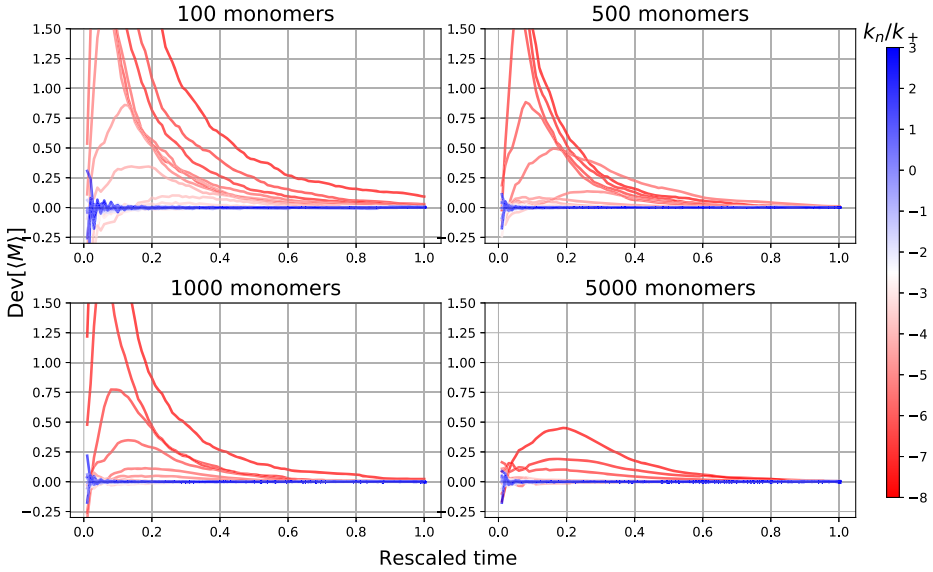
$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle P(t)m(t)\rangle &= k_n\langle m(t)^{n_c+1}\rangle - n_c k_n\langle P(t)m(t)^{n_c}\rangle \\ &\quad - n_c k_n\langle m(t)^{n_c}\rangle - 2k_+\langle P(t)^2 m(t)\rangle + 2k_-\langle P(t)^2\rangle \\ &\quad + k_f\langle m(t)(m_t - m(t)) - (2n_c - 1)P(t)m(t)\rangle \\ &\quad + k_f n_c(n_c - 1)\langle P(t)^2\rangle + k_2\langle m(t)^{n_2+1}(m_t - m(t))\rangle \\ &\quad - n_2 k_2\langle P(t)m(t)^{n_2}(m_t - m(t))\rangle \\ &\quad - n_2 k_2\langle m(t)^{n_2}(m_t - m(t))\rangle.\end{aligned} \tag{8}$$

For $n_c = 2$, there will be higher moments in eq 8, such as $\langle P(t)m(t)^2\rangle$. The moment closure technique then approximates those higher moments by using combinations of their lower moments,

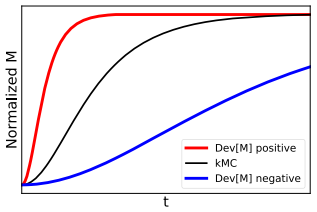
$$\begin{aligned}\langle P(t)m(t)^2\rangle &\simeq \frac{\langle P(t)m(t)\rangle^2\langle m(t)^2\rangle}{\langle P(t)\rangle\langle m(t)\rangle^2} \\ \langle P(t)^2 m(t)\rangle &\simeq \frac{\langle P(t)^2\rangle\langle P(t)m(t)\rangle^2}{\langle P(t)\rangle^2\langle m(t)\rangle} \\ \langle m(t)^3\rangle &\simeq \frac{\langle m(t)^2\rangle^3}{\langle m(t)\rangle^3}.\end{aligned} \tag{9}$$

Using the moment closure approximations shown in eq 9, the equations become a set of deterministic ordinary differential equations (ODEs), which can then be numerically solved to obtain both the mean and the variance of the fluctuations of the aggregation kinetics. The moment closure method makes solving eq 7 mathematically feasible. There is, however, a technical difficulty when resorting to numerical solutions on a computer. Since the approximation of the higher moments in terms of lower moments involves a large power of lower moments, the numerical solution of the equations can be prone to instability. There are two sources of "high power" terms: Hierarchy truncation effects and the intrinsically nonlinear terms that come from the nucleation parameters n_c and n_2. We of course need at least to calculate the variance of $P(t)$ and m, i.e., the second moments. Truncating the system at higher order results, however, results in a large value for the power of the moment that is used in approximations. This makes finding a stable numerical procedure challenging. We have found that using the truncation order set at two provides a high quality approximation and also that the solutions are stable.

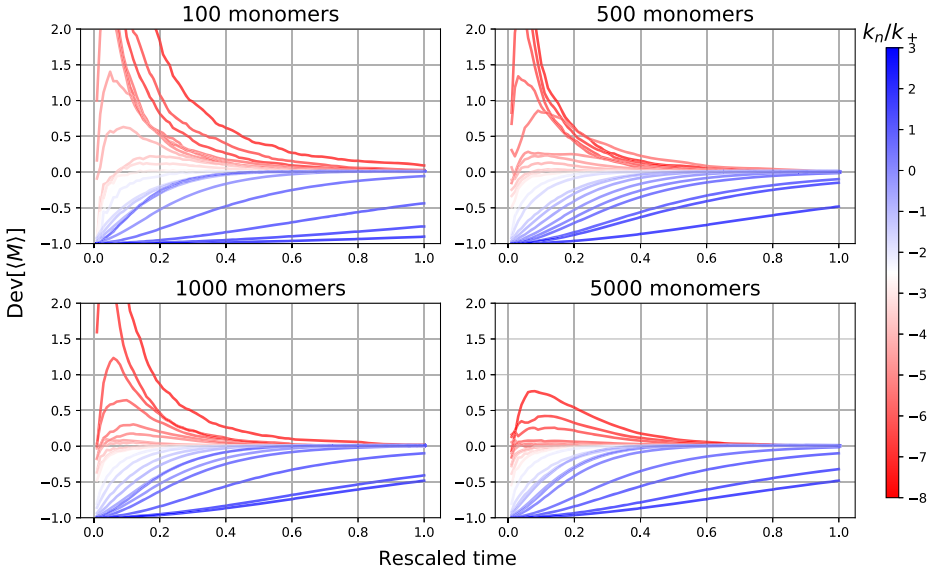
https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133



(a) Stochastic model with closeorder = 2



(b) Illustration of deviation



(c) Deterministic model

Figure 2. Deviations of $\langle M \rangle$ arising from the two approximations from the kMC results are plotted as a function of time (in a rescaled unit). Different colors represent the degree of deviation, with red and blue indicating positive and negative deviation, respectively. The corresponding kinetic profiles are illustrated in part b. (a) Stochastic model with close order set at 2. (c) Deterministic model. The deviations for different ratios of k_n/k_+ are represented by different colors and are plotted against a rescaled time axis, where the value of 1 on the time axis corresponds to the time when $\langle M \rangle$ reaches over 90% of the given number of monomers. The deviation is defined as $\mathrm{Dev}[\langle M \rangle] \equiv (\langle M \rangle_{\mathrm{model}} - \langle M \rangle_{\mathrm{kmc}})/\langle M \rangle_{\mathrm{kmc}}$. We held the elongation rate fixed at $k_+ = 1\ \mathrm{s}^{-1}$ and varied k_n. The following parameters are used: $P(0) = 0$, $n_c = 2$, $k_- = 10^{-6}\ \mathrm{s}^{-1}$, $k_f = 10^{-8}\ \mathrm{s}^{-1}$, and $k_2 = 0$.

The nucleation parameters n_c and n_2 determine the power of m that is found in the moment closures in eq 9. The resulting moment closure equations can become intractable if the value of n_c is large. We have however tested the solution numerically with n_c up to 4 and have found that, in any case, the approximation works reasonably well. We have found in contrast that the current moment closure approximation does not work well when n_2 is large, and that the numerical solutions then are not stable. To describe such a case will require alternative more appropriate approximations to handle the stochastic moment equations, shown in eq 7. We will leave this task for future work.

IV. RESULTS AND DISCUSSION

IV.A. The Averaged Fibril Mass $\langle M \rangle$. Before discussing fluctuations, we shall first calculate the average fibril mass, $\langle M(t) \rangle$, using both the deterministic equation and their stochastic form and compare results with the average fibril mass found in the benchmark kMC simulations. We vary the ratio of k_n to k_+ and carry out the analysis for systems with different total numbers of monomers. We have learned from the analysis that the ratio of k_n to k_+ (k_+ is held fixed at $1\ \mathrm{s}^{-1}$) controls the kinetic profile of the aggregation process, while the total number of monomers plays the dominant role in determining the extent of the fluctuations. The magnitude of

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

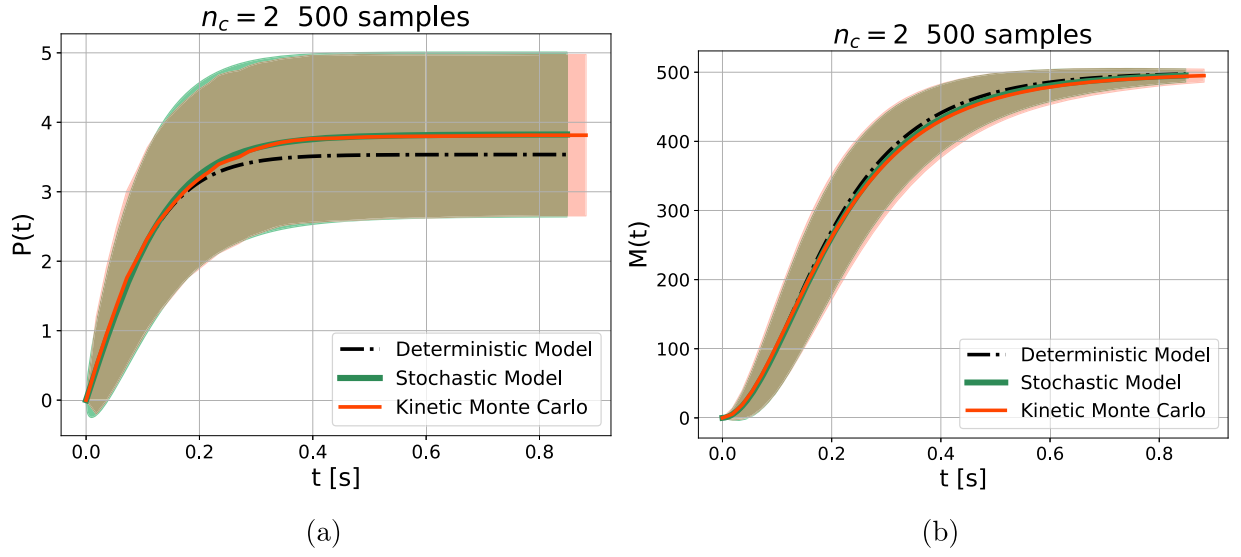


Figure 3. Kinetic profiles of $P(t)$ and $M(t)$ from the stochastic moment closure and from kinetic Monte Carlo simulations are shown and compared. (a) $P(t)$. (b) $M(t)$. The kinetic curve (mean and standard deviation) given by our stochastic model are represented by green solid lines and green shadows, respectively. The results from the kinetic MC (mean and standard deviation) are denoted by red solid lines and red shadows, respectively. The dashed black lines represent the results from the deterministic model. The parameter set of parts a and b that is used is identical with that of Figure 1.

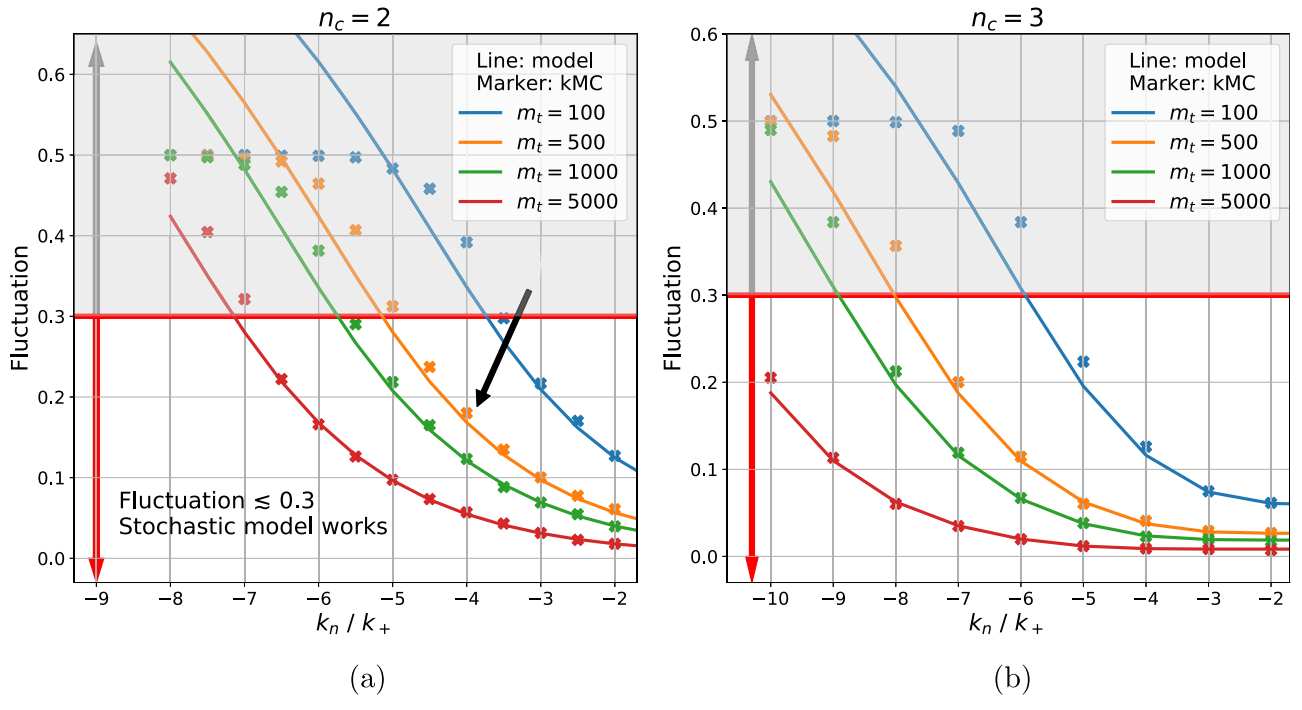


Figure 4. Quantitative analysis of the stochastic fluctuations of the mass concentration as a function of the kinetic ratio k_n/k_+ is compared for systems with different number of monomers. (a) $n_c = 2$. (b) $n_c = 3$. Each line represents the result from our stochastic model with a specified initial number of monomers, color-coded (blue, 100; orange, 500; green, 1000; red, 5000); the markers are the corresponding kMC simulation results. To distinguish the deviations of the model from the kMC results, a clear boundary (in red) is drawn at fluctuation = 0.3. Good agreement between the model and the kMC can be observed when the fluctuation is less than 0.3 (a red arrow pointing downward). Large deviation, however, can be seen when the fluctuation is larger than 0.3 (shaded in gray along with a gray arrow pointing upward). Note that the black arrow pointing to the data point shown in part a indicates the kinetic parameter set (m_t = 500) used in Figure 3.

the deviations of the $\langle M(t)\rangle$ obtained by second stochasticization from the profile found with kMC as a function of time (in reduced unit) is shown in Figure 2. The deviation is defined as $\mathrm{Dev}[\langle M\rangle] \equiv (\langle M\rangle_{\mathrm{model}} - \langle M\rangle_{\mathrm{kmc}})/\langle M\rangle_{\mathrm{kmc}}$. Each curve corresponds to a specific value of the ratio k_n/k_+, with red and blue indicating small and large value of k_n (it ranges from 10^{-8} to 10^{3}), respectively. Several features emerge in the analysis that are worth mentioning. First of all, the deviations are the largest for both models when k_n is extremely small ($<10^{-6}$) compared to k_+ (red in the color bar). As k_n increases, the deviations from the kMC simulation gradually decrease. Second, we observe that, for both models, the total number of

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

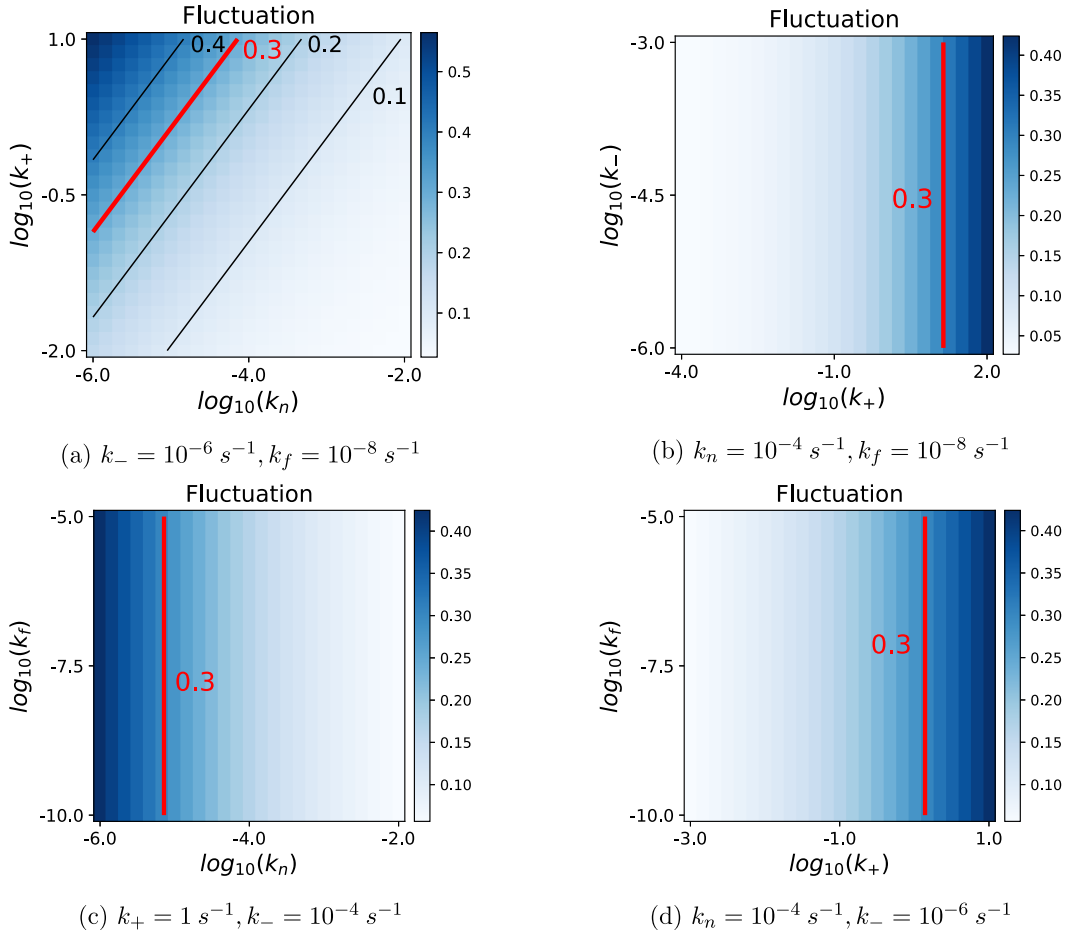


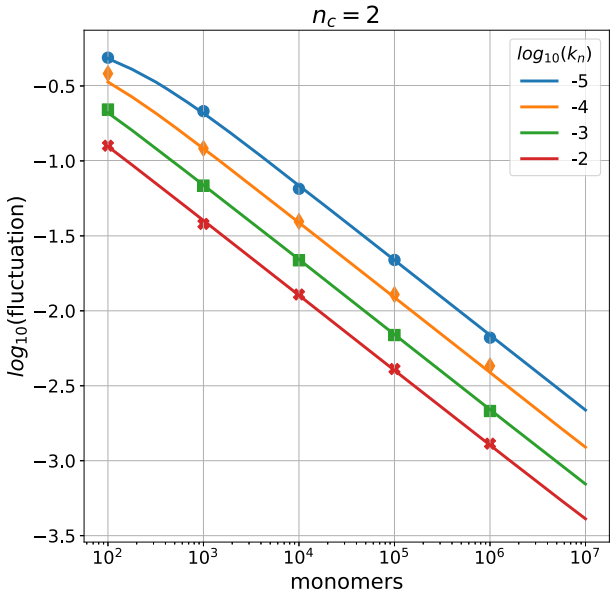
Figure 5. Fluctuation, Max(SD[M])/m_t, as a function of pairwise kinetic parameters (k_n, k_+, k_-, k_f) in log scale are shown. Only pairs with specific features are shown. (a) k_n vs k_+ (b) k_+ vs k_-(c) k_n vs k_f (d) k_+ vs k_f. The lines represent the value of the fluctuation, with red indicating 0.3. Note that for (a)(b)(c)(d) m_t = 500, $P(0)$ = 0, n_c = 2, k_2 = n_2 = 0, system closed at the second order. Rate constants are shown in s^{-1}.

monomers inversely correlates with the size of the deviations. As the total number of monomers increases from 100 to 5000, for example, the deviation decreases significantly. In general, the stochastic model based on the moment closure works well when $k_n > 10^{-3}$ while the deterministic approximation fails in this regime. The actual kinetic profiles for $\langle M(t)\rangle$ can be found in the Supporting Information.

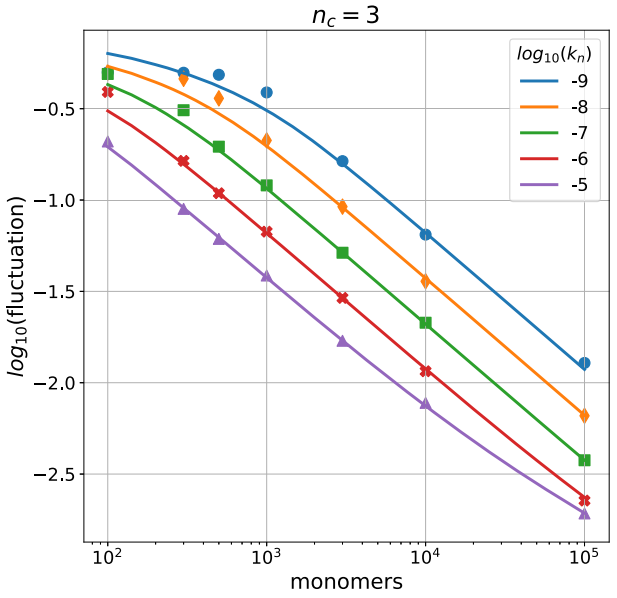
IV.B. Kinetic Profiles, *P*(*t*) and *M*(*t*), for Stochastic Aggregation. The time evolution of the mean and the standard deviations of $P(t)$ and $M(t)$ are easily obtained using second stochasticization. We compare the kinetic profiles of our stochastic model with those from the kMC simulations. The results of the two approaches are in good agreement with each other, as shown in Figure 3. To compare the stochastic fluctuations calculated from the fluctuating moment equations with those obtained from the kMC simulation, we use the maximum of the standard deviation of $M(t)$ throughout the time divided by the total initial number of monomers, yielding Max(SD[M])/m_t. We then carried out a parameter scan over k_n and m_t while keeping k_+ constant. The results are shown in Figure 4. Figure 4a shows the stochastic fluctuation as a function of the kinetic ratio k_n/k_+ for systems (n_c = 2) with different total numbers of monomers (m_t = 100, 500, 1000, 5000). Overall, the fluctuations increase as k_n decreases (or equivalently k_+ increases). At the same time, however, there is a greater discrepancy between the stochastic model and the kMC results as k_n decreases. In the small fluctuation region (fluctuation <0.3), the stochastic moment model in general agrees with the kMC results for the number of total monomers in the range of 100−5000. This result indicates there is a clear threshold for the fluctuation level where the deviations of the stochastic moment equation from the kMC simulation nearly can be considered negligible. When the fluctuations are large ($\gtrsim$ 0.3), however, the deviations between the two approaches become increasingly significant. The size of the fluctuations seems to show a linear dependence on k_n/k_+ when the ratio decreases, as does that from the kMC simulation. As k_n/k_+ continues to decrease, the size of the fluctuations finally saturates at a value (reaching a maximum of ≈0.5) due to the broad distribution of resulting lag times. This saturation phenomenon arises for an extreme situation where the mean value of $P(t)$ becomes very small ($\overline{P}(t) \leq 1$), which leads to a failure of the fluctuation analysis based on the total mass of aggregates. In this case, one can consider a different measure of stochasticity instead, for example, the lag time. Figure 4b presents another similar fluctuation view graph but now with n_c = 3.

IV.C. Parameter Space Scanning: How Variations of Different Rate Constants Affect the Fluctuation. In the previous section, we have shown that the model provides a good approximation for the kinetic profile when the fluctuations are well below a threshold value (Max(SD[M])/m_t < 0.3). Next, we shall investigate the effect of changing kinetic parameters on the stochasticity. This parameter

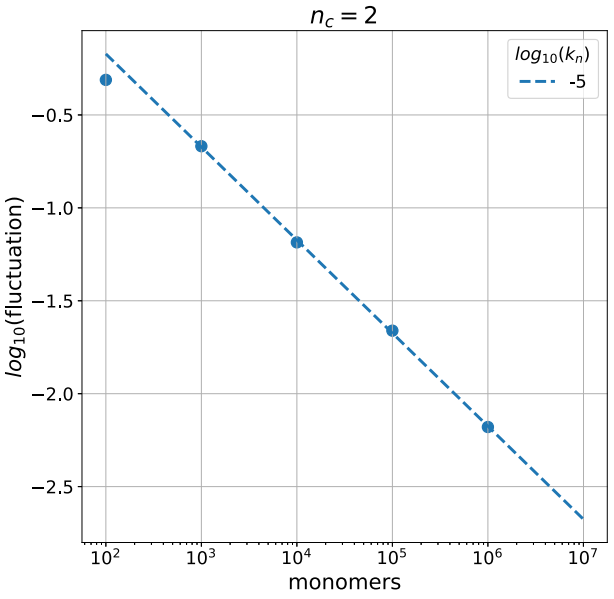
https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133



(a) $k_n = 10^{-5} - 10^{-2}$



(b) $k_n = 10^{-9} - 10^{-5}$



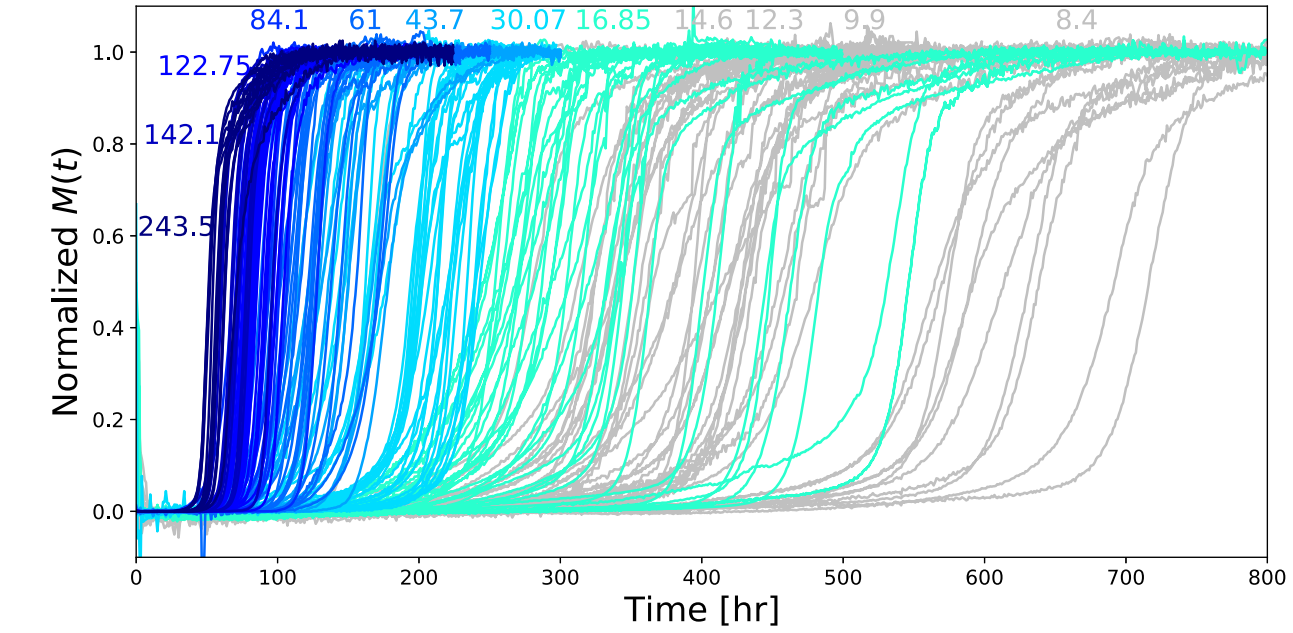
(c) $k_n = 10^{-5}$

Figure 6. The log−log plots of the fluctuation versus initial number of monomers m_t. (a) n_c = 2 (b) n_c = 3. The solid lines are the numerical solutions from our stochastic model, and the dots are the results from the kMC simulations. Different lines represent different k_n values, color-coded. In the case of part a (n_c = 2), the lines seem to share a common slope down the linear regime. In the case of part b (n_c = 3), despite some deviations found in low monomers regime, similar linear behavior can be identified. (c) The dashed line is used to approximate the linear relationship found in part a. The slope of the line is −0.5, which indeed corresponds to n_c = 2 according to eq 10. Note that k_+ = 1 s^{-1}, k_- = 10^{-6} s^{-1}, and k_f = 10^{-8} s^{-1} are used for both part a and part b.

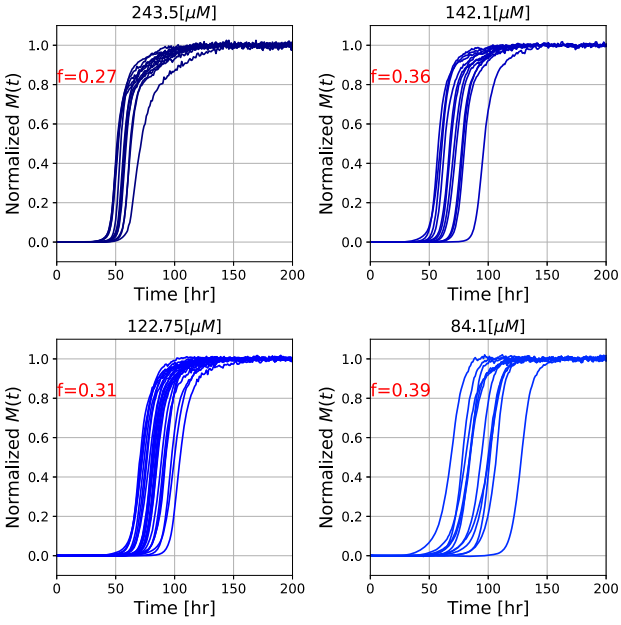
scanning survey provides a map of the size of stochastic fluctuations in relation to variations in the values of kinetic parameters. Figure 5 shows the maps of the size of the fluctuations. Different pairs of kinetic parameters are used: (k_+, k_n), (k_-, k_+), (k_f, k_n), and (k_f, k_+). These plots provide a visual guidance for the range of the parameter values involved. For example, the map of k_+ against k_n shown in Figure 5a shows that the model works well in the bottom right of the area defined by the red line (Max(SD[M])/m_t = 0.3). Other maps (Figure 5b and 5c) showing (k_-, k_+) and (k_f, k_n) pairs demonstrate that the model works well on the left and right-hand sides of the red line, respectively.

IV.D. A Scaling Relation. Fluctuations are influenced by the ratio of the kinetic parameters, k_n/k_+, suggesting there is a

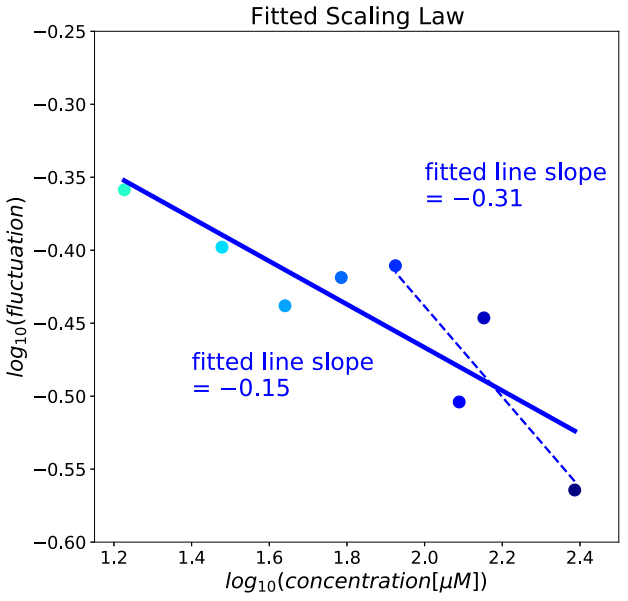
https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133



(a) Kinetic traces for a range of concentrations



(b) Selected concentrations



(c) Scaling behavior

Figure 7. A verification of the scaling law found in the numerical simulations using experimental data is shown. (a) Time evolution of normalized M from experiments of different concentrations, ranging from 8.4 to 243.5 μM. (b) Selected kinetic traces for the medium-to-high concentration regime. Note that the number in the red text indicates the size of fluctuations. See Supporting Information for the kinetic traces for the full concentration regime. (c) Fitted scaling behavior of the fluctuations over the experimental concentrations (in log $_{10}$ scale). The fitted slope for 8 data points ($\geq$16.85 μM) with a wide range of fluctuations ($0.27 \leq f \leq 0.45$) is −0.15 (blue solid line); whereas for large concentrations ($\geq$84.1 μM; 4 independent kinetic traces shown in part b) with a smaller upper bound of fluctuations ($0.27 \leq f \leq 0.40$), the slope is −0.31 (blue dashed line). Using the scaling law we show in the previous section, the observed slope gives n_c, which is $0.15 \times 4 = 0.6$ and $0.31 \times 4 = 1.24$, respectively, for both cases. Data and figure reproduced with permission from ref 28. Copyright 2008 National Academy of Science.

generic scaling behavior that correlates with the system size. Apparently, there is a scaling relation between the fluctuations of $M(t)$ with the total number of monomers m_t which correlates with the critical nucleus size, at least up to $n_c = 4$. Similar to the analysis described above, we quantify the fluctuations using the maximum of standard deviation of $M(t)$ throughout the time, Max(SD[$M(t)$]). The magnitude of the fluctuations as a function of total number of monomers m_t, both in log scale, is shown in Figure 6. The scaling law therefore reads

$$\text{Fluctuation} := \text{Max}(\text{SD}[M(t)]) \sim m_t^{-n_c/4} \tag{10}$$

From the log−log plot of the fluctuations versus the total number of monomers m_t, we can see how the scaling law

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

changes with k_n. For different values of k_n we see that the lines share a common slope, and the slope is indeed determined by n_c (slope = $-n_c/4$). The results for $n_c = 2$ and $n_c = 3$ are shown in parts a and b of Figure 6, respectively. In the case $n_c = 2$, for example, the slope is −0.5, which indeed yields $n_c = 2$. This result supports the validity of the scaling relation. Interestingly, we observed a mild deviation from the linearity when m_t is small ($\approx 10^2$). This "bending" phenomenon is notable for $n_c = 3$, in particular when primary nucleation becomes slow ($k_n < 10^{-6}$).

IV.E. Scaling in Experiments. The experiments of Xue et al. provide a set of data that record the time evolution of the mass of aggregates (or the fibril mass) starting from different initial protein concentrations that range from few μM to about 240 μM.[28] The data show a variation of the fibril mass against protein concentration. The exact mathematical relationship is expected to follow the scaling law shown in eq 10. To facilitate the analysis, we manually divided the data into several groups each with a sufficient number of samples; these groups were then labeled with approximated centered concentrations from which the standard deviations can be calculated.

Figure 7 shows that the experimental data do confirm the scaling law. From the data, we have extracted the mean values of the protein concentrations and the corresponding deviations from the mean. We then looked at how the size of the fluctuations varies with concentration. Figure 7a shows multiple kinetic traces of β_2 microglobulin aggregation at different protein concentrations (8.4−243.5 μM). Indeed, as the concentration decreases, the kinetic traces become more dispersed (shown using curves in different color scales). In the fitting procedure, the data in low concentration regime (<16 μM, in grayscale colors) were not used due to their extremely large fluctuation. The data used therefore can be divided into two data sets according to concentration range: Set I, 16 μM ≤ data ≤ 243 μM (8 kinetic traces); Set II, 80 μM ≤ data ≤ 243 μM (4 kinetic traces out of Set I, that is Set II ⊆ Set I). For each of the concentrations, the size of fluctuations is calculated; the value ranges from 0.27 (highest concentration) to 0.45 (lowest concentration). The overall fluctuation−concentration relation follows the trend: the more concentrated the protein is in solution, the smaller is the size of fluctuations. Figure 7b shows the kinetic traces at four specific concentrations of Set II (84.1, 122.75, 142.1, and 243.5 μM) as examples. The size of fluctuations of individual kinetic traces in Set I are shown against their concentration, presented in Figure 7c. The size of the fluctuations follows the linear relationship, complying with the scaling law. From the linear fit for Set I, the slope is −0.15 (solid blue line), which yields the value of n_c = 0.6 whereas the slope for Set II is −0.31 (dashed blue line) with n_c = 1.24. The latter prediction provides a somewhat steeper slope in high concentration regime (84 to 243 μM). This result suggests that the effective critical nucleus size in this regime is somewhere between 1 (monomeric nucleus) and 2 (dimeric nucleus), consistent with the existing interpretations for the growth nucleus size.[7] We notice, however, that the definition we have used for the size of the nuclei is somewhat different from that used in the experimental paper.[28] Xue et al. also discussed a generalized mechanistic model for fibril assembly process. Here, we do not employ the same definitions used by Xue et al.

V. CONCLUSIONS

In this work, we have developed a new approach to study the stochastic aspects of protein aggregation in living cells, where the number of protein copies is usually low as compared to experiments used in *in vitro* studies. Using the chemical Langevin approach, we have developed a mathematical moment closure technique that arrives at a set of stochastic differential equations with very fewer variables than in standard Monte Carlo models. The resulting stochastic moment equations take into account nevertheless the stochastic noise based on the kinetic propensities (i.e., rate constants) of individual chemical reactions. The approach allows efficient stochastic analysis as well as accurate predictions for averaged kinetic profiles when the fluctuations are not too large. We have compared this "second stochasticized" model with the deterministic one. The stochastic model performs better than the deterministic model in the description of fibril mass concentration $\langle M(t)\rangle$, particularly in the fast nucleation regime ($k_n/k_+ > 10^{-3}$). In addition, quantitative analysis of the stochastic fluctuations of the fibril mass concentration shows that the model successfully describes stochastic fluctuation of $M(t)$ even at quite low protein copy number (m_t can be as small as 10^2) when $k_n/k_+ > 10^{-3}$, which suggests the second stochasticization approach can be usefully applied in the cell, where usually only a small number of protein copies are available ($m_t \sim 10^2$). We also have found a physical scaling law that correlates the total number of protein copies with the magnitude of the stochastic fluctuations. We have verified the scaling relation using experimental data in the medium-to-high concentration regime (>16 μM). The result agrees with the inferred critical nucleus size, suggesting that the proposed model, along with its universality and the derived scaling law, will be generally useful for studying the stochasticity of protein aggregation. Conceptually similar to the scaling relation, a somewhat different but interesting relation that correlates the rates of nucleation via stochasticity was noted some decades ago. Eaton and Hofrichter found that when secondary nucleation is present, the rate of primary nucleation can be obtained from the distribution of the lag time.[29] Their result clearly pointed out the role of stochasticity in inferring mechanistic details, i.e., nucleation processes.

Although we have seen that second stochasticization allows a quantitative description of stochasticity of protein aggregation in well mixed systems, we would like to point out a conceptual limitation of the present calculations. The "lag time" is a characteristic time scale of aggregation. The localized character of nucleation implies that spatial variations of the specific concentrations and thus their moments are nearly certain to be important in real cells. We note however that the second stochasticization approximation can be extended to deal with spatial variations using stochastic partial differential equations. We leave that extension to future work.

APPENDIX: CALCULATION OF THE NOISE TERMS IN THE STOCHASTIC MOMENT EQUATIONS

In this section, we show how to obtain the stochastic moment equations, eq 7, by summing over the noises in the stochastic rate equation, eq 4. The calculation of the deterministic part in eq 7 from eq 4 can be found in literature.[8] We will focus on the summation over the noises here.

In the stochastic rate equation, we have labeled the noises with the associated chemical reactions. The noises are

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

independent when they arise from different chemical reactions, while the noises of the same type of chemical events are identical and thus they can cancel out each other. It is important to remember the case that the differences between two independent noises are generally nonzero. This correlation is thus a constraint needed to obey the conservation of the number of monomers. We will deal with the noises in elongation first, and then the noises in fragmentation. Calculations for the other terms follow the same construction.

We first calculate the noises of elongation in the $P(t)$ moment equation. Summing over j on the $P(t)$ moment equation, we have

$$\begin{aligned}
&\sum_{j=n_c}^{\infty} \sqrt{2k_+m(t)}\,[\sqrt{f(j-1)}\,\xi(A_{j-1}+A_1 \to A_j) - \sqrt{f(j)}\,\xi(A_j + A_1 \to A_{j+1})] \\
&= \sqrt{2k_+m(t)}\left[\sum_{j=n_c+1}^{\infty} \sqrt{f(j-1)}\,\xi(A_{j-1}+A_1 \to A_j) - \sum_{j=n_c}^{\infty} \sqrt{f(j)}\,\xi(A_j + A_1 \to A_{j+1})\right] \\
&= \sqrt{2k_+m(t)}\left[\sum_{j=n_c}^{\infty} \sqrt{f(j)}\,\xi(A_j + A_1 \to A_{j+1}) - \sum_{j=n_c}^{\infty} \sqrt{f(j)}\,\xi(A_j + A_1 \to A_{j+1})\right] = 0,
\end{aligned} \tag{11}$$

where we have used the fact that $f(j < n_c) = 0$, and we relabeled the dummy index j in the third line. The noises in the first and second term are identical since they originate from the same chemical reaction, $A_j + A_1 \to A_{j+1}$. Thus, they could not be treated as independent noises and the difference between them is zero.

The noises of elongation in the $M(t)$ moment equation are (ignoring the common factor $\sqrt{2k_+m(t)}$)

$$\begin{aligned}
&\sum_{j=n_c}^{\infty} j[\sqrt{f(j-1)}\,\xi(A_{j-1}+A_1 \to A_j) - \sqrt{f(j)}\,\xi(A_j + A_1 \to A_{j+1})] = \sum_{j=n_c+1}^{\infty} j\sqrt{f(j-1)}\,\xi(A_{j-1}+A_1 \to A_j) \\
&- \sum_{j=n_c}^{\infty} j\sqrt{f(j)}\,\xi(A_j + A_1 \to A_{j+1}) = \sum_{j=n_c}^{\infty} \sqrt{f(j)}\,\xi(A_j + A_1 \to A_{j+1}) = \sqrt{\sum_{j=n_c}^{\infty} f(j)} \times \xi = \sqrt{P(t)}\,\xi,
\end{aligned} \tag{12}$$

where $\xi \sim \mathcal{N}(0, 1)$. In eq 12, we have used two properties of Gaussian random variables. First, if $\xi \sim \mathcal{N}(0, 1)$ then $c\xi \sim \mathcal{N}(0, c^2)$. Second, if ξ and ξ' are independent random variables and $\xi \sim \mathcal{N}(0, {\sigma_\xi}^2)$, $\xi' \sim \mathcal{N}(0, {\sigma_{\xi'}}^2)$ then $(\xi + \xi') \sim \mathcal{N}(0, {\sigma_\xi}^2 + {\sigma_{\xi'}}^2)$.

The noises for the fragmentation reaction in eq 4 are summations over several different fragmentation noises, because there are many channels to get or to dissolve a certain length of aggregate through the fragmentation reaction. To calculate the moment of $f(t, j)$, we have to sum over all $j \geq n_c$, and so we have terms involving double summations. We can exchange the order of the summation and relabel the dummy index i and j (we will use $\xi[a, b, c]$ to represent $\xi(A_a \to A_b + A_c)$),

$$\sum_{j=n_c}^{\infty}\sum_{i=j+1}^{\infty} \sqrt{f(i)}\,\xi[i, j, i-j] = \sum_{i=n_c+1}^{\infty}\sum_{j=n_c}^{i-1} \sqrt{f(i)}\,\xi[i, j, i-j] = \sum_{j=n_c+1}^{\infty} \sqrt{f(j)} \sum_{i=n_c}^{j-1} \xi[j, i, j-i]. \tag{13}$$

Now, this expression above can be combined with the other fragmentation noises in eq 4, since they both have a summation over j from $n_c + 1$ to the infinity in front of them, and we get

$$\begin{aligned}
\frac{\mathrm{d}P(t)}{\mathrm{d}t} &= \sum_{j=n_c}^{\infty}\sum_{i=j+1}^{\infty} \sqrt{2k_f f(j)}\,\xi[i, j, i-j] + (2-\sqrt{2}) \sum_{j=n_c}^{\infty} \sqrt{k_f f(2j)}\,\xi[2j, j, j] \\
&+ \begin{cases}
-\sum_{j=n_c+1}^{\infty} \sqrt{2k_f f(j)} \sum_{i=1}^{(j-1)/2} \xi[j, j-i, i] & j \text{ odd} \\
-\sum_{j=n_c+1}^{\infty} \sqrt{2k_f f(j)} \sum_{i=1}^{(j-2)/2} \xi[j, j-i, i] - \sum_{j=n_c+1}^{\infty} \sqrt{k_f f(j)}\,\xi\left[j, \frac{j}{2}, \frac{j}{2}\right] & j \text{ even}
\end{cases}
\end{aligned}$$

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

$$
\begin{aligned}
&=(\sqrt{2}-1)\sum_{j=n_c}^{\infty}\sqrt{2k_f f(2j)}\,\xi[2j,j,j]\\
&\quad+\sum_{j=n_c+1}^{\infty}\sqrt{2k_f f(j)}\left[\sum_{i=n_c}^{j-1}\xi[j,i,j-i]+\begin{cases}-\sum_{i=1}^{(j-1)/2}\xi[j,j-i,i] & j\ \text{odd}\\ -\sum_{i=1}^{(j-2)/2}\xi[j,j-i,i]-\frac{1}{\sqrt{2}}\xi\left[j,\frac{j}{2},\frac{j}{2}\right] & j\ \text{even}\end{cases}\right]\\
&=\sum_{j=n_c+1}^{\infty}\sqrt{2k_f f(j)}\left[\sum_{i=1}^{j-n_c}\xi[j,j-i,i]+\begin{cases}-\sum_{i=1}^{(j-1)/2}\xi[j,j-i,i] & j\ \text{odd}\\ -\sum_{i=1}^{(j-2)/2}\xi[j,j-i,i]+\left(\frac{\sqrt{2}}{2}-1\right)\xi\left[j,\frac{j}{2},\frac{j}{2}\right] & j\ \text{even}\end{cases}\right]\\
&=\sum_{j=n_c+1}^{\infty}\sqrt{2k_f f(j)}\begin{cases}\sum_{i=(j+1)/2}^{j-n_c}\xi[j,i,j-i] & j\ \text{odd}\\ \sum_{i=j/2}^{j-n_c}\xi[j,i,j-i]+\left(\frac{\sqrt{2}}{2}-1\right)\xi\left[j,\frac{j}{2},\frac{j}{2}\right] & j\ \text{even}\end{cases}\\
&=\sum_{j=n_c+1}^{\infty}\sqrt{2k_f f(j)}\begin{cases}\sum_{i=(j+1)/2}^{j-n_c}\xi[j,i,j-i] & j\ \text{odd}\\ \sum_{i=j/2+1}^{j-n_c}\xi[j,i,j-i]+\left(\frac{\sqrt{2}}{2}\right)\xi\left[j,\frac{j}{2},\frac{j}{2}\right] & j\ \text{even}\end{cases}
\end{aligned}
\tag{14}
$$

So far, we have not used the properties of the Gaussian white noise. The sum of Gaussian white noises in the odd j case is a new Gaussian white noise with a variance $(j-n_c)-(j+1)/2+1=(j+1-2n_c)/2$, whereas in the even j case, it is $(j-n_c)-(j/2+1)+1+(1/2)=(j+1-2n_c)/2$. The variances for both cases are the same, so we can use a single expression,

$$
\begin{aligned}
\frac{\mathrm{d}P(t)}{\mathrm{d}t}&=\sum_{j=n_c+1}^{\infty}\sqrt{2k_f f(j)}\,\xi_j, \qquad \xi_j\sim\mathcal{N}\left(0,\frac{j+1-2n_c}{2}\right)\\
&=\sqrt{\sum_{j=n_c+1}^{\infty}k_f(j+1-2n_c)f(j)}\,\xi_{fP}, \qquad \xi_{fP}\sim\mathcal{N}(0,1)\\
&=\sqrt{k_f[M-(2n_c-1)P(t)]}\,\xi_{fP},
\end{aligned}
\tag{15}
$$

where we have used the fact that all the ξ_j are mutually independent, since they are from different fragmentation channels, and this allows us to write a new Gaussian white noise ξ_{fP}.

To calculate the fragmentation noise in the $M(t)$ equation, we first rearrange the double sum in the first fragmentation line in eq 4,

$$
\begin{aligned}
&\sum_{j=n_c}^{\infty}j\sum_{i=j+1}^{\infty}\sqrt{f(i)}\,\xi[i,j,i-j]=\sum_{j=n_c}^{\infty}\sum_{i=j+1}^{\infty}j\sqrt{f(i)}\,\xi[i,j,i-j]=\sum_{i=n_c+1}^{\infty}\sum_{j=n_c}^{i-1}j\sqrt{f(i)}\,\xi[i,j,i-j]\\
&\quad=\sum_{j=n_c+1}^{\infty}\sqrt{f(j)}\sum_{i=n_c}^{j-1}i\xi[j,i,j-i].
\end{aligned}
\tag{16}
$$

Then, we have

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

$$\frac{dM(t)}{dt} = \sum_{j=n_c}^{\infty} j \sum_{i=j+1}^{\infty} \sqrt{2k_f f(j)}\,\xi[i, j, i-j] + (2-\sqrt{2}) \sum_{j=n_c}^{\infty} j\sqrt{k_f f(2j)}\,\xi[2j, j, j]$$
$$+ \begin{cases} -\sum_{j=n_c+1}^{\infty} j\sqrt{2k_f f(j)} \sum_{i=1}^{(j-1)/2} \xi[j, j-i, i] & j \text{ odd} \\ -\sum_{j=n_c+1}^{\infty} j\sqrt{2k_f f(j)} \sum_{i=1}^{(j-2)/2} \xi[j, j-i, i] - \sum_{j=n_c+1}^{\infty} j\sqrt{k_f f(j)}\,\xi\left[j, \frac{j}{2}, \frac{j}{2}\right] & j \text{ even} \end{cases}$$
$$= (\sqrt{2}-1) \sum_{j=n_c}^{\infty} j\sqrt{2k_f f(2j)}\,\xi[2j, j, j] + \sum_{j=n_c+1}^{\infty} \sqrt{2k_f f(j)} \left[\sum_{i=n_c}^{j-1} i\xi[j, i, j-i] \right.$$
$$\left. + \begin{cases} -\sum_{i=1}^{(j-1)/2} j\xi[j, j-i, i] & j \text{ odd} \\ -\sum_{i=1}^{(j-2)/2} j\xi[j, j-i, i] - \frac{j}{\sqrt{2}}\xi\left[j, \frac{j}{2}, \frac{j}{2}\right] & j \text{ even} \end{cases} \right] \tag{17}$$

By inspection, we find that the above equation, for both cases, can be written

$$\frac{\mathrm{d}M(t)}{\mathrm{d}t} = -\sum_{j=n_c+1}^{\infty} \sqrt{2k_f f(j)} \sum_{i=1}^{n_c-1} i\xi[j, i, j-i] = -\sum_{j=n_c+1}^{\infty} \sqrt{2k_f f(j)} \sum_{i=1}^{n_c-1} \xi_{[j,i]}, \qquad \xi_{[j,i]} \sim \mathcal{N}(0, i^2)$$
$$= -\sum_{j=n_c+1}^{\infty} \sqrt{2k_f f(j)}\,\xi_j, \qquad \xi_j \sim \mathcal{N}\left(0, \sum_{i=1}^{n_c-1} i^2\right) = \mathcal{N}\left(0, \frac{n_c(n_c-1)(2n_c-1)}{6}\right)$$
$$= -\sqrt{\sum_{j=n_c+1}^{\infty} k_f f(j) \frac{n_c(n_c-1)(2n_c-1)}{3}}\,\xi_{fM}, \qquad \xi_{fM} \sim \mathcal{N}(0, 1)$$
$$= -\sqrt{k_f \frac{n_c(n_c-1)(2n_c-1)}{3} P(t)}\,\xi_{fM}. \tag{18}$$

Combining the results from above, we obtain the stochastic moment equation in eq 7.

We have also used Mathematica's symbolic calculation to verify the above results, by explicitly listing all the resulting noise terms after the addition or the subtraction. We use different labels for the fragmentation noises in thed $P(t)/\mathrm{d}t$ and $\mathrm{d}M(t)/\mathrm{d}t$ equations. In fact, a careful comparison between ξ_{fP} and ξ_{fM} shows that, interestingly, they are *almost* independent. The summed over fragmentation channels in the two cases have only tiny intersection, resulting from the cases when fragmentation reactions change both the number of aggregates and the total mass in aggregates. These are rare events, since this kind of fragmentation can only occur when an aggregate with length $n_c < l < 2n_c$ breaks, resulting in less one in $P(t)$ and less l in $M(t)$ (the shorter than n_c aggregate is assumed to dissolve into monomers immediately after being produced). Note that this independence is exact when $n_c = 2$. Usually, in a protein aggregation system, the averaged size of aggregates is much larger than n_c, for the typical cases with n_c being small; therefore, the assumption that the two noises ξ_{fP} and ξ_{fM} are independent is justified. When calculating the population moment $\langle P^\alpha M^\beta \rangle$, we will set $\langle \xi_{fP}\xi_{fM} \rangle = 0$.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c10331.

Figure S1, kinetic profiles $\langle M(t) \rangle$ for different k_n/k_+ ratios ($m_t = 500$); Figure S2, $d\langle M(t) \rangle/dt$ as a function of time for different k_n/k_+ ratios ($m_t = 500$); Figure S3, examples of kinetic profiles of $P(t)$ and $M(t)$ showing large deviations from the kMC results ($m_t = 500$); Figure S4, experimental kinetic traces of β_2 microglobulin aggregation at different protein concentrations from 8.4 to 243.5 μM (PDF)

https://dx.doi.org/10.1021/acs.jpcb.0c10331
J. Phys. Chem. B 2021, 125, 1118−1133

AUTHOR INFORMATION

Corresponding Authors

Min-Yeh Tsai − *Department of Chemistry, Tamkang University, New Taipei City 251301, Taiwan;* orcid.org/0000-0002-6275-0312; Email: mytsai@mail.tku.edu.tw

Peter G. Wolynes − *Center for Theoretical Biological Physics and Department of Chemistry, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-7975-9287; Email: pwolynes@rice.edu

Authors

Jia-Liang Shen − *Department of Chemistry, Tamkang University, New Taipei City 251301, Taiwan*

Nicholas P. Schafer − *Center for Theoretical Biological Physics and Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c10331

Author Contributions

[§]J.-L.S. and M.-Y.T. contributed equally to this work

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Center for Theoretical Biological Physics sponsored by the National Science Foundation (USA), Grant PHY-2019745. Additional support was provided by the D. R. Bullard-Welch Chair at Rice University, Grant C-0016, and the Ministry of Science and Technology (MOST), Taiwan (R.O.C.), Grant No. 108-2113-M-032-003-MY2.

REFERENCES

(1) Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426*, 884.

(2) Murphy, M. P.; LeVine, H. Alzheimer's Disease and the Beta-Amyloid Peptide. *J. Alzheimer's Dis.* **2010**, *19*, 311.

(3) Oosawa, F.; Kasai, M. A theory of linear and helical aggregations of macromolecules. *J. Mol. Biol.* **1962**, *4*, 10−21.

(4) Oosawa, F.; Asakura, S. *Thermodynamics of the Polymerisation of Proteins*; Academic Press: London and New York, 1975.

(5) Ferrone, F. A.; Hofrichter, J.; Eaton, W. A. Kinetics of sickle hemoglobin polymerization: II. A double nucleation mechanism. *J. Mol. Biol.* **1985**, *183*, 611−631.

(6) Bishop, M. F.; Ferrone, F. A. Kinetics of nucleation-controlled polymerization. A perturbation treatment for use with a secondary pathway. *Biophys. J.* **1984**, *46*, 631−644.

(7) Knowles, T. P. J.; Waudby, C. A.; Devlin, G. L.; Cohen, S. I. A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* **2009**, *326*, 1533−1537.

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

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

(10) Cohen, S. I. A.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. From macroscopic measurements to microscopic mechanisms of protein aggregation. *J. Mol. Biol.* **2012**, *421*, 160−171.

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

(12) Szavits-Nossan, J.; Eden, K.; Morris, R. J.; MacPhee, C. E.; Evans, M. R.; Allen, R. J. Inherent Variability in the Kinetics of Autocatalytic Protein Self-Assembly. *Phys. Rev. Lett.* **2014**, *113*, 098101.

(13) Michaels, T. C. T.; Dear, A. J.; Kirkegaard, J. B.; Saar, K. L.; Weitz, D. A.; Knowles, T. P. J. Fluctuations in the Kinetics of Linear Protein Self-Assembly. *Phys. Rev. Lett.* **2016**, *116*, 258103.

(14) Michaels, T. C. T.; Dear, A. J.; Knowles, T. P. J. Stochastic calculus of protein filament formation under spatial confinement. *New J. Phys.* **2018**, *20*, 055007.

(15) Gillespie, D. T. The chemical Langevin equation. *J. Chem. Phys.* **2000**, *113*, 297−306.

(16) Lu, T.; Hasty, J.; Wolynes, P. G. Effective Temperature in Stochastic Kinetics and Gene Networks. *Biophys. J.* **2006**, *91*, 84−94.

(17) Brett, T.; Galla, T. Stochastic Processes with Distributed Delays: Chemical Langevin Equation and Linear-Noise Approximation. *Phys. Rev. Lett.* **2013**, *110*, 250601.

(18) Ghusinga, K. R.; Soltani, M.; Lamperski, A.; Dhople, S. V.; Singh, A. Approximate moment dynamics for polynomial and trigonometric stochastic systems. *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. 2017; pp 1864−1869.

(19) Szabo, A. Fluctuations in the polymerization of sickle hemoglobin. A simple analytic model. *J. Mol. Biol.* **1988**, *199*, 539−542.

(20) Sanft, K. R.; Othmer, H. G. Constant-complexity stochastic simulation algorithm with optimal binning. *J. Chem. Phys.* **2015**, *143*, 074108.

(21) Michaels, T. C. T.; Garcia, G. A.; Knowles, T. P. J. Asymptotic solutions of the Oosawa model for the length distribution of biofilaments. *J. Chem. Phys.* **2014**, *140*, 194906.

(22) Michaels, T. C. T.; Knowles, T. P. J. Mean-field master equation formalism for biofilament growth. *Am. J. Phys.* **2014**, *82*, 476−483.

(23) Michaels, T. C. T.; Cohen, S. I. A.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Hamiltonian Dynamics of Protein Filament Formation. *Phys. Rev. Lett.* **2016**, *116*, 038101.

(24) Cohen, S. I. A.; Linse, S.; Luheshi, L. M.; Hellstrand, E.; White, D. A.; Rajah, L.; Otzen, D. E.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Proliferation of amyloid-Beta42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 9758−9763.

(25) Chiti, F.; Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86*, 27−68.

(26) Schreck, J. S.; Yuan, J.-M. A kinetic study of amyloid formation: fibril growth and length distributions. *J. Phys. Chem. B* **2013**, *117*, 6574−6583.

(27) Taniguchi, Y.; Choi, P. J.; Li, G.-W.; Chen, H.; Babu, M.; Hearn, J.; Emili, A.; Xie, X. S. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **2010**, *329*, 533−538.

(28) Xue, W.-F.; Homans, S. W.; Radford, S. E. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 8926−8931.

(29) Eaton, W. A.; Hofrichter, J. Sickle cell hemoglobin polymerization. *Adv. Protein Chem.* **1990**, *40*, 63−279.