Insights into the variability of nucleated amyloid polymerization by a minimalistic model of stochastic protein assembly

Sarah Eugène*
INRIA de Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France and Sorbonne Universités, UPMC Université Pierre et Marie Curie, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

Wei-Feng Xue†
School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, UK

Philippe Robert‡
INRIA de Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France

Marie Doumic§
INRIA de Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France and Sorbonne Universités, UPMC Université Pierre et Marie Curie, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

(Dated: April 8, 2016)

Self-assembly of proteins into amyloid aggregates is an important biological phenomenon associated with human diseases such as Alzheimer's disease. Amyloid fibrils also have potential applications in nano-engineering of biomaterials. The kinetics of amyloid assembly show an exponential growth phase preceded by a lag phase, variable in duration as seen in bulk experiments and experiments that mimic the small volumes of cells. Here, to investigate the origins and the properties of the observed variability in the lag phase of amyloid assembly currently not accounted for by deterministic nucleation dependent mechanisms, we formulate a new stochastic minimal model that is capable of describing the characteristics of amyloid growth curves despite its simplicity. We then solve the stochastic differential equations of our model and give mathematical proof of a central limit theorem for the sample growth trajectories of the nucleated aggregation process. These results give an asymptotic description for our simple model, from which closed form analytical results capable of describing and predicting the variability of nucleated amyloid assembly were derived. We also demonstrate the application of our results to inform experiments in a conceptually friendly and clear fashion. Our model offers a new perspective and paves the way for a new and efficient approach on extracting vital information regarding the key initial events of amyloid formation.

INTRODUCTION

The amyloid conformation of proteins is of increasing concern in our society because they are associated with devastating human diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, Prion diseases and type-2 diabetes [1, 2]. The fibrillar assemblies of amyloid are also of considerable interest in nanoscience and engineering due to their distinct functional and materials properties [3–6]. Elucidating the molecular mechanism of how proteins polymerize to form amyloid oligomers, aggregates and fibrils is, therefore, a fundamental challenge for current medical and nanomaterials research.

Amyloid diseases are associated with the aggregation and deposition of mis-folded proteins in the amyloid conformation [1, 2]. Amyloid aggregates form through nucleated polymerization of monomeric protein or peptide precursors (e.g. [7–11]). The slow nucleation process that initiates the conversion of proteins into their amyloid conformation is followed by exponential growth of amyloid particles, resulting in growth of amyloid fibrils that is accelerated by secondary processes such as fibril fragmentation and aggregate surface catalyzed heterogeneous nucleation [7, 11–13] (Figure 1). Current mathematical description of protein assembly into amyloid are based on systems of mass-action ordinary differential equations, and they have been successful in describing the average behaviour of amyloid assembly observed by kinetic experiments (e.g.[11, 12]). The formation kinetics of amyloid aggregates has been studied extensively by bulk *in vitro* experiments in volumes typically in the range of hundreds of μL or larger [12], but has also been observed recently in elegant microfluidic experiments in pL to nL range, more closely mimicking physiological volumes in tissues and cellular compartments [14]. Amyloid growth experiments typically follow the appearance of amyloid

* Sarah.Eugene@inria.fr
† W.F.Xue@kent.ac.uk; http://www.kent.ac.uk/bio/profiles/staff/xue.html
‡ Philippe.Robert@inria.fr, corresponding author; http://team.inria.fr/rap/robert
§ Marie.Doumic@inria.fr; https://www.rocq.inria.fr/bang/Marie-Doumic
PR and MD contributed equally to supervise this work.

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

aggregates or the depletion of monomers as function of time, yielding information regarding the rate of the exponential growth and the length of the lag phase under different protein concentrations at fixed volumes. A hitherto overlooked piece of information that can be derived from these kinetic experiments is the observed variation between experimental repeats, which may hold the key to understanding the early rare nucleation events of amyloid formation [12, 15–17]. However, current deterministic models cannot describe variability, thus, unable to address whether the observed variations in lag phase length reflect subtle experimental differences between the replicates, contributions from the stochastic nature of the nucleation mechanism, or a combination of both factors. As shown recently by Szavits-Nossan and co-workers using a stochastic nucleated growth model, rare nucleation events are expected to dictate the behaviour and variability of amyloid formation in small volumes such as in cellular compartments [15]. Understanding these rare initial nucleation events of amyloid formation and the variability resulting from the stochastic nature of nucleation, therefore, is of paramount importance in the fundamental understanding of amyloid diseases and in controlling amyloid formation.

Here, we present a new stochastic protein assembly model with the aim to capture the fundamental features of amyloid self-assembly that includes their stochastic nature, and still allow a fully rigorous mathematical analysis of these processes (Figure 1). In this spirit, our model contains minimal possible complexity needed to describe a nucleated protein polymerization process, allowing us to study it theoretically in a mathematically rigorous manner, but still allowing useful comparison to experimental data. From our minimal model, we derive a closed form formula that can describe and predict variability in the lag phase duration of nucleated protein assembly by giving a proof to a central limit theorem for our model. Our results demonstrate how stochasticity at the molecular level may influence the kinetics of the total reaction population at a macroscopic scale depending on the relative rates of nucleation and exponential growth, and on reaction volume. We also show how new information relevant to any specific nucleated amyloid assembly can be gained in a conceptually simple and clear manner by applying our analytical results to the analysis of published $\beta_2 m$ amyloid assembly kinetics data [12]. We demonstrate that our model qualitatively captures key features of the data such as parallel progress of the curves and the order of magnitude for the rates of the self-accelerating reactions. We also show that the intrinsic stochastic nature of nucleation alone cannot explain the observed variability in lag phase length for published $\beta_2 m$ amyloid assembly data acquired in large (100 μL) volumes suggesting alternative mechanistic assembly steps and additional experimental sources that contribute to the lag-time variability in the observed amyloid growth curves. Our approach represents the basis for the development of extensive and tractable stochastic models, which will allow the variability information from amyloid growth kinetics experiments to be used to inform the fundamental molecular mechanisms of the key rare initial events of amyloid formation that may be involved in producing early on-pathway cytotoxic species associated with amyloid disease.

See supplemental material [18] at [URL will be inserted by AIP] for the mathematical background of these results, in particular the rigorous proofs of the convergence results, the precise mathematical characterization of the variability of the assembly process and, finally, some simulations of these stochastic processes.

FIG. 1. (A) Schematic illustration of a full amyloid assembly model, including conformational exchange, nucleation, elongation growth, secondary surface nucleation and fibril fragmentation processes. (B) Schematic illustration of the minimal model represented by reactions (1) and (2). The phenomenological parameters α and β represent the rate constant of the ignition phase, and the rate constant of all possible accelerated growth pathways to the formation of polymers, respectively. The circles represent the un-polymerised monomer $\mathcal{X}$ and the parallelograms represent the monomeric units in the amyloid formation $\mathcal{Y}$ in (1) and (2). Some monomeric units are highlighted with bold outlines to highlight few possible paths a monomeric unit in (1) and (2) can take through the aggregation process.

A PHENOMENOLOGICAL STOCHASTIC MODEL

To make the model as simple as possible, we consider two distinct types of monomers, we call these species monomers ($\mathcal{X}$) and polymerised monomers ($\mathcal{Y}$), respectively. The polymerised monomers $\mathcal{Y}$ represent all monomeric units in the amyloid conformation in the aggregates. Its amount may be viewed as representing the total polymerised mass, captured for instance by Thioflavine T (ThT) measurements, as example shows in Figure 2. Such a simplification is also justified by the fact that current kinetics measurements of amyloid growth exhibit variability on the timecourse of the total polymerised mass, without giving any information on the size distribution of fibrils. However, such a simplification do not contain any contributions from spatial dynamics, molecular motion and transport processes, which may add complexity to the stochastic behaviour in small volumes. Previous studies (see for instance [19], Supplemental material (S.M.) 2) have shown that the detail of the reactions of secondary pathways, such as a fragmentation kernel, may have a major impact on the size distribution of polymers, but comparably smaller effects on the timecourse of the polymerised mass. Overall, with this simplification, we can distill the problem down from infinite number of species to two species, which subsequently can describe the ability of the amyloid state to convert normal un-polymerised monomers to the amyloid state with-

out invoking polymer ends or number. Thus, our model is phenomenological and aims to give new insights into the key determinants of stochastic behavior of protein aggregation and suggests simple ways to extract information from experimental data. Our approach departs from the mechanistic modelling approach used in conventional deterministic models of protein aggregation but is complementary to those models (e.g. [11, 12, 15]), and the simplifications allows tractable mathematical derivation of closed expressions.

We thus consider two distinct species in our model: monomers, $\mathcal{X}$, and polymerised monomers, $\mathcal{Y}$. We then consider $X_V(t)$ and $Y_V(t)$ to be the respective numbers of particles of each species at time t in a fixed volume V. Initially, it is assumed that there are only M monomers: $X_V(0) = M$ and $Y_V(0) = 0$. We denote $m = M/(V{\cdot}N_A)$ the initial molar concentration of monomers, where N_A is the Avogadro constant. For convenience in the calculations hereinafter, we introduce the notation $V_A = V{\cdot}N_A$.

Thus, the chemical reactions associated with this simple model are as follows:

$$\mathcal{X} + \mathcal{X} \xrightarrow{\alpha/V_A^2} 2\mathcal{Y}, \tag{1}$$

$$\mathcal{X} + \mathcal{Y} \xrightarrow{\beta/V_A^2} 2\mathcal{Y}, \tag{2}$$

where α/V_A^2 and β/V_A^2 are rates of the reactions with rate constants of $\alpha > 0$ and $\beta > 0$. These reactions describe the following features of a nucleated polymerisation of proteins that characterises amyloid assembly (see Figure 1 for an explanatory scheme of the reactions):

— Reaction (1): We call this step "ignition" since it models the starting point of the polymerisation process. Here, we represents this step as the simplest possible concentration dependent nucleation step that converts two monomers into two monomeric units in the amyloid state that are growth competent (equivalent to two polymerised monomers). The initiation step (1) is equivalent to a nucleation step involving dimer formation. This is a common simplification that has been applied in a number of deterministic model (e.g. [11, 13]), and is also motivated by the fact that the first molecular attachment step towards the nucleation barrier may have the biggest energetic penalty according to the classical nucleation theory. In our model, this reaction will occur in a stochastic way. Following the principles of the law of mass action, the encounter of two chemical species occurs at a rate proportional to the product of the *concentrations* of each species. Therefore two monomers ($\mathcal{X}$) disappear to produce two polymerised monomers ($\mathcal{Y}$) at a rate α/V^2.

— Reaction (2): We call this second step "conversion", which we modelled as a self-accelerating autocatalytic process. Here, given a monomer and a polymerised monomer, the monomer converts into a polymerised monomer at a rate β/V^2. This is representative of a range of accelerating secondary pathway reactions such as fragmentation, lateral growth, and aggregate surface catalyzed second nucleation. In this sense, our model may be viewed as a phenomenological simplification and amalgamation of several mechanistic models. Even though different secondary processes lead to very different size distributions of fibrils, they affect the total polymerised mass, represented here by the quantity of the species $\mathcal{Y}$, in a qualitatively similar way in that they provide acceleration of growth through positive feedback. For example, in the case of fragmentation accelerated growth, fibril fragments that interact with monomers $\mathcal{X}$ are generated, in a first order approximation, proportional to the number of breakage sites [7], which in-turn depends on the number of monomeric units in the amyloid fibrils. In the case of secondary fibril surface nucleation, the sites that promote surface nucleation is proportional to available surface [13], which is also dependent on the number of monomeric units in the amyloid fibrils (Fig 1). In particular, we expect our model to behave qualitatively similarly to the mechanistic model described in [15], which includes nucleation, polymerization, and fragmentation as a representative self-accelerating secondary process motivated by its experimental analysis [14]. It is however not intended for reaction (2) to be associated to any specific microscopic meaning as described above.

FIG. 2. Examples of experimental timecourses of amyloid assembly reactions [12]. Twelve experimental timecourse of polymerised mass for two given initial concentrations of monomers are shown: 122 μM (blue) and 30.5 μM (green).

Stochastic Evolution. Any given pair of monomers reacts together by Reaction (1) at rate α/V_A^2, whereas for a given pair of monomer/polymerized monomer reacts by Reaction (2) at rate β/V_A^2. Let M_V be the initial number of monomers and the random variable describing the number of monomers remaining at time t is denoted by $X_V(t)$. By taking into account the $X_V(X_V{-}1)/2$ monomers pairs, and the $X_V(M_V{-}X_V)$ monomers/polymerised pairs, the variable $X_V(t)$ has jumps of size -2 or -1 which occur at the following rates

$$X_V \mapsto \begin{cases} X_V{-}2 & \text{at rate } \dfrac{X_V(X_V{-}1)}{2} \times \dfrac{\alpha}{V^2}, \\ X_V{-}1 & \text{“} \quad X_V(M_V{-}X_V) \times \dfrac{\beta}{V^2}. \end{cases} \tag{3}$$

The conservation of mass gives the additional relation $X_V(t){+}Y_V(t){=}M_V$. As noticed previously, in the description of Reactions (1) and (2) above, this representation is completely coherent with the law of mass action.

A. Asymptotic Evolution of the Number of Monomers

Assuming that the volume V is large and the initial concentration of monomers remains constant and equal to $m > 0$, i.e. the initial number of monomers M_V is such that $M_V/V_A \sim m$, we can derive the following:

Polymerisation occurs on the time scale $t \mapsto V_A t$. Let $(\overline{X}_V(t))$ be the scaled process defined by

$$\overline{X}_V(t) = \frac{X_V(V_A t)}{V_A}. \tag{4}$$

In Equation (4), the time scale of the process $(X_V(t))$ is accelerated with a factor V_A. As it will be seen, as V gets large, $t \to V_A t$ is the correct time scale to carry out a large volume expansion and to observe the decay of $(X_V(t))$ on the space scale proportional to V_A.

Assuming for the moment that $(\overline{X}_V(t))$ is converging in distribution, Relations (3) then suggest that its limit $(\bar{x}(t))$ should satisfy the following ODE, which is the law of Mass Action for our simple model

$$\frac{\mathrm{d}\bar{x}}{\mathrm{d}t} = -\alpha\bar{x}(t)^2 - \beta\bar{x}(t)(m - \bar{x}(t)), \text{ with } \bar{x}(0) = m. \tag{5}$$

One can show that this is indeed the case. If the initial number M_V of monomers is such that

$$\lim_{V\to+\infty} \frac{M_V}{V_A} = m > 0,$$

then, as V goes to infinity, the process $(\overline{X}_V(t))$ converges in distribution to $(\bar{x}(t))$, solution of Equation (5), given by the formula

$$\bar{x}(t) = m\frac{\beta}{\alpha}\frac{1}{e^{\beta m t} - 1 + \beta/\alpha}. \tag{6}$$

The proof is classical [20], we recall it in Sections A and E of supplemental material, we comment on the relative influence of the parameters α and β on the deterministic curve, see supplemental figure S2. This asymptotic result is illustrated with simulations in Figure 3.

In order to be able to quantify the variability of experimental replicates, we need to go further, to a second order approximation, i.e. with a central limit result. If the initial number M_V of monomers is such that

$$M_V = mV_A + o\left(\sqrt{V_A}\right),$$

for $m > 0$, then, for the convergence in distribution,

$$\lim_{V\to+\infty} \left(\frac{X_V(V_A t) - V_A\bar{x}(t)}{m\sqrt{V_A}}\right) = (U(t)),$$

where $U(t)$ is a diffusion, the unique solution of the following stochastic differential equation (S3).

The proof is postponed in Section B of supplemental material, together with an explicit formulation and an analysis of the influence of the parameters α and β on the stochasticity of the reactions. We found that the smaller the ratio α/β is, the more important the influence of the stochasticity on the lag-time, but the less important for the following of the reaction. This is quantified in the following study of the stochastic time for δ completion below.

FIG. 3. Simulations of the model with different numerical parameter values. Parameters and the time axes are scaled with identical units. The red curve is the first order solution of Equation(5).

B. Asymptotics of the Time for δ Reaction Completion

To quantify the effect of α and β on the stochasticity of the reactions, we define the *time for δ reaction completion*, where $0 < \delta < 1$ is a percentage, as the following stopping time

$$T_V(\delta) = \inf\{t > 0, X_V(t) \le (1-\delta)M_V\}$$

where T_V is the first time when there is a δ fraction of polymers is produced. T_V for δ small - 5 to 10% - represents an alternative definition for the lag-time of the reaction[19].

A law of large numbers and a central limit result for $T_V(\delta)$ as V goes to infinity can be obtained. Note that due to the change in the time scale, we need to rescale T_V by V to get a limit.

If the initial number M_V of monomers is such that

$$M_V = mV_A + o\left(\sqrt{V_A}\right),$$

for $m > 0$ then, for the convergence in distribution

1. **Law of Large Numbers**.

$$\lim_{V\to+\infty} \frac{T_V(\delta)}{V_A} = t_\delta \overset{\text{def.}}{=} \frac{1}{\beta m}\log\left(1 + \frac{\beta\delta}{\alpha(1-\delta)}\right). \tag{7}$$

2. **Central Limit Result**.

$$\lim_{V\to+\infty} \frac{T_V(\delta) - V_A t_\delta}{\sqrt{V_A}} = \frac{U(t_\delta)}{m[\alpha(1-\delta)^2 + \beta\delta(1-\delta)]}$$

where $(U(t))$ is the solution of the SDE (S3).

The proof of this result is given in Section C of supplemental material. Fig. S1 illustrates and the law of large numbers and the central limit theorem for $T_{1/2}$. Note that the definition of t_δ, which is the limit of the stochastic times $T_V(\delta)/V$ when $V \to \infty$ is coherent with the definition of the *deterministic* time of δ reaction completion as

$$t_\delta = \inf\{t > 0, \bar{x}(t) \le (1-\delta)m\} = \bar{x}^{-1}((1-\delta)m),$$

where $(\bar{x}(t))$ is given by Equation (6). Thus, for any given experiment, the distance between a realization of $T_V(\delta)$ and t_δ is being given by the explicit formula above. We can therefore derive its stochastic behaviour. Under the assumptions of the above theorem and with its notations, the variance σ_V^2 of the time for δ completion has a limit σ^2, when $\alpha \ll \beta$

$$\lim_{V\to+\infty} \sigma_V = \sigma \sim \frac{\sqrt{3}}{\sqrt{2}m\sqrt{M_V\alpha\beta}}. \tag{8}$$

The proof is given in Section D of supplemental material, together with the exact formula (S9) for σ of supplemental material. Interestingly, this result obtained from our minimal model is comparable to the expression on lag-time variations obtained in [15] based on a more complex mechanistic model by the mean of a Taylor expansion. This result, therefore, corroborates with the idea that our minimum model with only ignition and conversion contains the key features sufficient in qualitatively describing the stochastic properties of the nucleated protein aggregation processes. Our simplified formula (8) and its full general form in the equation (S9) of the supplemental information, are mathematically fully rigorous, and allows analysis of the intricate interplay between the ignition reaction and the autocatalytic reaction. In fact, it is possible to have a whole range of times when both reactions have an influence over the whole aggregation timecourse, as may be seen on Formula (8).

It should be noted that this representation of σ is independent of δ. This suggests that the fluctuations do not depend on δ, and therefore, the growth curves predicted by our simple model are all parallel for any given concentration. Figures 5 (c) and 5 (d) below have been obtained by centering the 12 curves of Figures 5 (a) and 5 (b) at the half-time corresponding to $\delta = 1/2$. As it can be seen, the times $T_V(\delta)$ for $0.4 \leq \delta \leq 0.7$ are then also superimposed: the curves are identical for this range of values. This is an illustration of the above relation (8). The exact mathematical formulation of this phenomenon is shown in supplemental material. Simple as it is, our model captures well this feature experimentally observed. Also, it emphasizes the fact that we can take different values for δ without having an influence on the study. A difficulty however lies in the fact that when the numerical values of the constant α above is in the order of $1/M_V$, then the convergence itself may be a problem, as it can be seen on Figure 4.

FIG. 4. Comparison between the simulations and the predictions to see the regime of parameters where the calculations are valid. For these simulations, we fixed $M_V = 10^7$ monomers, $\beta = 1$, and made α varying.

C. Estimation of the parameters

In this section, we tested our theoretical results obtained with our minimalistic stochastic model on the experimental data published in [12]. Our model suggests a straightforward manner by which experimental amyloid aggregation timecourse curves can be analysed to give information on the expected variability of the lag-time length resulting from the stochastic nature of nucleated polymerisation. More precisely, using Formula (5) for k and Relation (7) for $t_{1/2}$, gives

$$t_{1/2} = \log\left(1+\beta/\alpha\right)/\beta m \text{ and } k = m\beta\left(1+\alpha/\beta\right)/4. \tag{9}$$

In the experiments in [12], there are 12 replicate kinetic traces reported for each sample concentration in constant 100L reaction volume. The parameters α and β can be obtained in a straightforward manner by fitting equations (9) to the mean half-time $t_{1/2}$ and the mean slope k of the curves at $t_{1/2}$. Table I shows a summary of our analysis. The constants α and β, and the calculated variance (8) are shown for each of the concentrations used. We also carried out a global analysis for α and β, fitting (9) simultaneously all of the curves for all concentrations. See Fig. S4. The overlay of the experimental curves around the predicted mean is illustrated in Figure 5, Figures 5 (c) and 5 (d) have been obtained by centering the 12 curves of Figures 5 (a) and 5 (b) at the half-time corresponding to $\delta = 1/2$. As can be seen, the agreement between the calculated and the experimental curves is good for $0.4 \leq \delta \leq 0.7$. This is consistent with the relation (8).

FIG. 5. (a) and (b): Experimental timecourse of polymerised mass for 12 different experiments [12]. (c) and (d): with a centering at the $T_V(1/2)$ of each curve. The red curve is the predicted mean with the estimated parameters.

Our analysis further demonstrates two important insights. Firstly, we obtained a more well-estimated β parameter. It is remarkable that the numerical value of β, which quantifies the conversion step in our model, does not change much for the 15 concentrations tested in the experiments, considering the simplicity of our model. This is not the case for α,which quantifies the ignition phase, varies between 10^{-7} and 10^{-13}. Here, the parameter α which quantifies the take-off phase (remember that the slope of $(\bar{x}(t))$ at 0 is $-\alpha m^2$) is intrinsically estimated with less precision than β, see Section E of supplemental material. This is a limitation of this simple model, and it also reflect the lack of information content in the kinetics data during the lag phase compared to the growth phase.

Secondly, despite good agreement between our analysis and the data in terms of the shapes of the growth curves, the analysis results in a much smaller order of magnitude for the variability among curves compared with experimental data. Since the relation $\alpha \ll \beta$

$\mathbf{m(10^{-6}M)}$	$\alpha(\mathbf{h^{-1}.M^{-1}})$	$\beta(\mathbf{h^{-1}.M^{-1}})$	**Exp. Dev.** (h)	**Predicted Dev.** (h)
12.3	$6.18{\cdot}10^{-7}$	$5.07{\cdot}10^{4}$	7.95	$5.34{\cdot}10^{-2}$
14.6	$2.81{\cdot}10^{-6}$	$4.54{\cdot}10^{4}$	2.98	$2.05{\cdot}10^{-2}$
16.7	$1.59{\cdot}10^{-4}$	$3.75{\cdot}10^{4}$	2.68	$2.45{\cdot}10^{-3}$
17.0	$1.88{\cdot}10^{-3}$	$3.70{\cdot}10^{4}$	1.52	$6.98{\cdot}10^{-4}$
29.5	$1.40{\cdot}10^{-5}$	$3.34{\cdot}10^{4}$	2.13	$3.7{\cdot}10^{-3}$
30.2	$2.89{\cdot}10^{-2}$	$2.96{\cdot}10^{4}$	2.57	$8.40{\cdot}10^{-5}$
30.5	$9.57{\cdot}10^{-8}$	$4.16{\cdot}10^{4}$	1.53	$3.84{\cdot}10^{-2}$
43.7	$7.99{\cdot}10^{-3}$	$2.35{\cdot}10^{4}$	2.10	$1.03{\cdot}10^{-4}$
48.5	$1.68{\cdot}10^{-2}$	$2.01{\cdot}10^{4}$	1.56	$6.55{\cdot}10^{-5}$
61.0	$2.61{\cdot}10^{-2}$	$2.04{\cdot}10^{4}$	1.03	$3.71{\cdot}10^{-5}$
61.0	$2.22{\cdot}10^{-5}$	$2.56{\cdot}10^{4}$	2.55	$1.14{\cdot}10^{-3}$
84.1	$4.53{\cdot}10^{-4}$	$2.24{\cdot}10^{4}$	1.59	$1.66{\cdot}10^{-4}$
102.2	$1.52{\cdot}10^{-3}$	$1.88{\cdot}10^{4}$	0.62	$7.39{\cdot}10^{-5}$
122	$1.33{\cdot}10^{-4}$	$1.75{\cdot}10^{4}$	0.90	$1.98{\cdot}10^{-4}$
123.5	$2.13{\cdot}10^{-4}$	$1.79{\cdot}10^{4}$	0.90	$1.52{\cdot}10^{-4}$
142.1	$2.58{\cdot}10^{-4}$	$1.74{\cdot}10^{4}$	1.11	$1.13{\cdot}10^{-4}$
243.5	$1.75{\cdot}10^{-3}$	$1.09{\cdot}10^{4}$	0.60	$2.46{\cdot}10^{-5}$

TABLE I. Parameters estimated from experiments [12] using our model. The two first columns are the estimated parameters α and β from the model. The third column is the experimental standard deviation of $T_V(1/2)$, while the fourth is the standard deviation predicted by our mathematical results for the model with the estimated parameters. We see that the estimation for β is quite robust, in contrast with that of α.

holds in the numerical estimations, Equation (8) gives the approximation $\sigma^2 \sim 3/(2M_V m^2 \alpha\beta)$ for the variance of the characteristic times of the kinetic traces. A variance of the order of magnitude observed in the experiments [12] would be obtained by our model for an initial number of monomers M_V in the order of 10^6. As the number M_V in the experiments of [12] performed in $100\mu L$ volumes is closer to 10^{15}, our analysis suggest that the variability observed result from more than a simple stochastic homogeneous nucleation of monomers. This result is consistent with the mechanistic approach used by Szavits-Nossan and co-workers [15], where the authors used a stochastic nucleation-polymerization-fragmentation based model. Thus, our model and analysis of the variance suggest alternative initial rare assembly steps that involve additional complexities such as conformational exchange, and/or additional experimental sources that contribute to the variability in the observed amyloid growth curves.

CONCLUSION

In this study, we described an approach that represents the molecular mechanisms of amyloid growth in a condensed way to enable the development of a rigorous framework that can describe the stochastic behaviour in addition to the general features of the kinetics of assembly. We adopted a reductionist approach by including minimal complexity using two simple processes, ignition and conversion, to reflect the idea that a protein polymerization reaction accelerated by secondary processes is dictated by the first nucleation event [14, 15]. With our minimal model, we were able to derive an exact analytical formula for the expected variability among curves as function of relative rates of the ignition and conversion processes, and the reaction volume. This is useful both in exploring the interplay between the reaction rates of nucleation and growth, and the variability in reaction traces. Our approach also suggest a straightforward manner in which information regarding the stochastic behaviour of nucleated protein aggregation can be extracted from experimental kinetics data using equations (8) and (9). We see that the stochasticity influences mainly the ignition step: once the reaction accelerates in the conversion step, all curves become parallel and deterministic, as illustrated both by experiments and the model we presented here. Thus, simple as it is, our model captures well the features experimentally observed for amyloid growth curves. Also, it confirms, as expected, that we can take different characteristic times (such as lag time, or growth mid point) when analysing kinetic growth curves. Our model further informed the need for new mechanistic steps or experimental interpretation of the large observed variations in the lag time lengths. Thus, the variation seen in the kinetic traces must be taken into account in addition to the concentration dependent behaviour of the kinetic traces in evaluating and developing mechanistic understanding of amyloid protein assembly processes.

While our model design was not aimed at describing the reality of any specific amyloid forming system with all of their individual associated complexities, our design by pursuing maximum simplicity are complementary to mechanistic approaches such as in [11, 12, 15] in capturing global properties of amyloid assembly. A particularly interesting direction for future work would be to envisage other orders for the reactions, in particular β, which cur-

rently is not specific to any particular accelerating growth processes, or an extended model with an initial conformational exchange step, for example. In summary, our current method allows for a rigorous theoretical treatment and understanding, and therefore, provides a basis for future model selection on stochastic minimal models, each of these models being the condensation of a family of possible stochastic mechanistic models that are closer to reality but for which analytical formulae are out of reach.

It should be remarked that in many cases the reaction curve shows significant asymmetries about the half-time which cannot be captured without accounting for the correct dependencies of the secondary nucleation rate on the monomer. Such an asymmetry is present e.g. in the data of Figures 2 and 5, take off is slower than the approach to the plateau, a common characteristic of systems dominated by fragmentation. However, the model predictions, e.g. in Figure 3, give curves that are perfectly symmetric about the half time (because of the term $X(M-X)$ in the rate equations). This suggests the possibility of different dependencies of the autocatalytic part on the monomer which we plan to investigate in the future.

ACKNOWLEDGMENTS

This research was supported by ERC Starting Grant SKIPPER^{AD} No. 306321 (M. Doumic and S. Eugène), and BBSRC grant BB/M02427X/1 (Wei-Feng Xue).

[1] T. P. Knowles, M. Vendruscolo, and C. M. Dobson, Nat Rev Mol Cell Biol **15**, 384 (2014).

[2] F. Chiti and C. M. Dobson, Annu Rev Biochem **75**, 333 (2006).

[3] D. M. Fowler, A. V. Koulov, W. E. Balch, and J. W. Kelly, Trends Biochem Sci **32**, 217 (2007).

[4] K. Schwartz and B. R. Boles, Curr Opin Microbiol **16**, 93 (2013).

[5] G. Meisl, J. B. Kirkegaard, P. Arosio, T. C. T. Michaels, M. Vendruscolo, C. M. Dobson, S. Linse, and T. P. J. Knowles, Nature Protocols **11**, 252 (2016).

[6] T. P. Knowles and M. J. Buehler, Nat Nanotechnol **6**, 469 (2011).

[7] W.-F. Xue and S. E. Radford, Biophys J **105**, 2811 (2013).

[8] D. Kashchiev and S. Auer, J Chem Phys **132**, 215101 (2010).

[9] F. A. Ferrone, Methods Enzymol **309**, 256 (1999).

[10] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman, PLoS Biol **2**, e321 (2004).

[11] T. P. Knowles, C. A. Waudby, G. L. Devlin, S. I. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, Science **326**, 1533 (2009).

[12] W.-F. Xue, S. W. Homans, and S. E. Radford, PNAS **105**, 8926 (2008).

[13] S. I. Cohen, S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson, and T. P. Knowles, Proc Natl Acad Sci U S A **110**, 9758 (2013).

[14] T. P. Knowles, D. A. White, A. R. Abate, J. J. Agresti, S. I. Cohen, R. A. Sperling, E. J. D. Genst, C. M. Dobson, and D. A. Weitz, Proc Natl Acad Sci U S A **108**, 14746 (2011).

[15] J. Szavits-Nossan, K. Eden, R. J. Morris, C. E. MacPhee, M. R. Evans, and R. J. Allen, Physical Review Letters **113**, 098101 (2014).

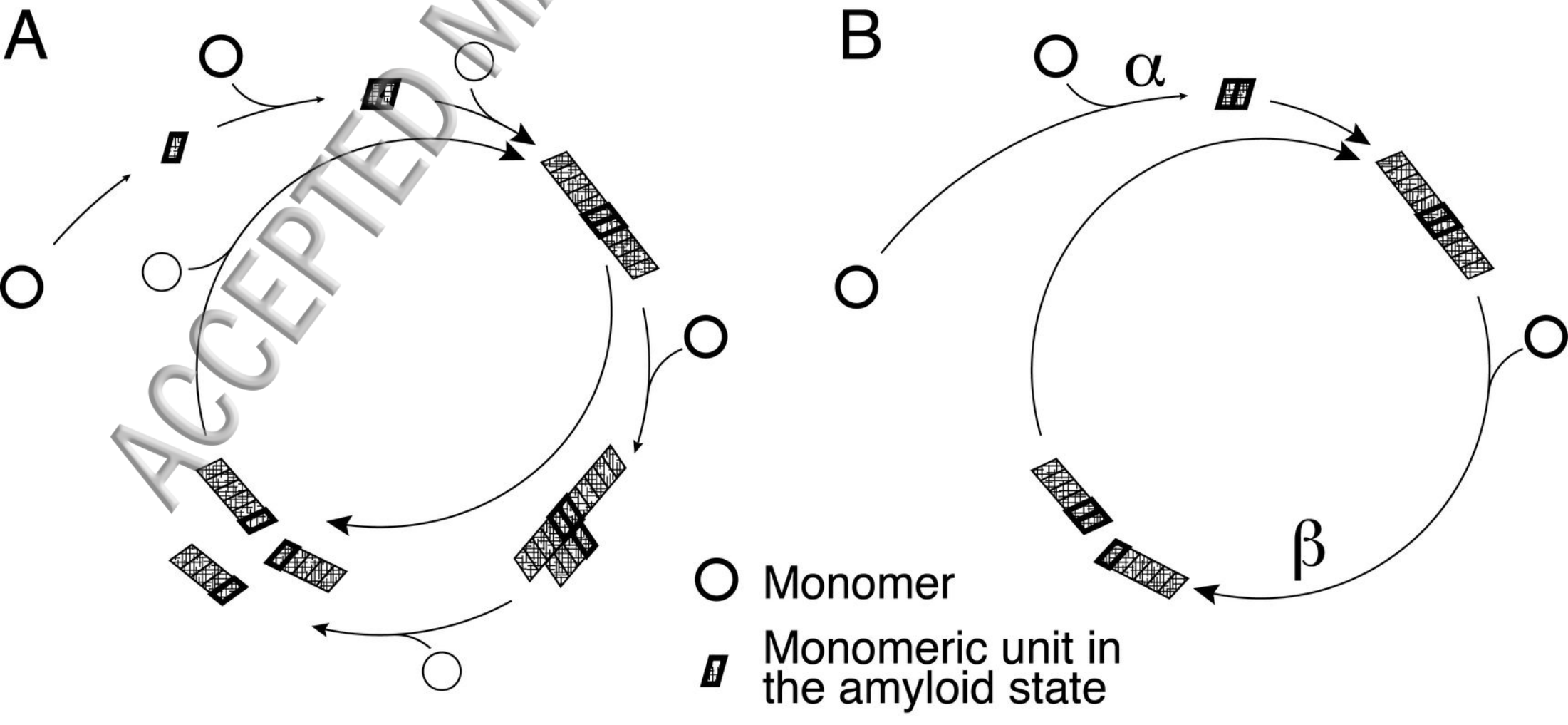
[16] F. A. Ferrone, J. Hofrichter, and W. A. Eaton, Journal of Molecular Biology **183**, 591 (1985).

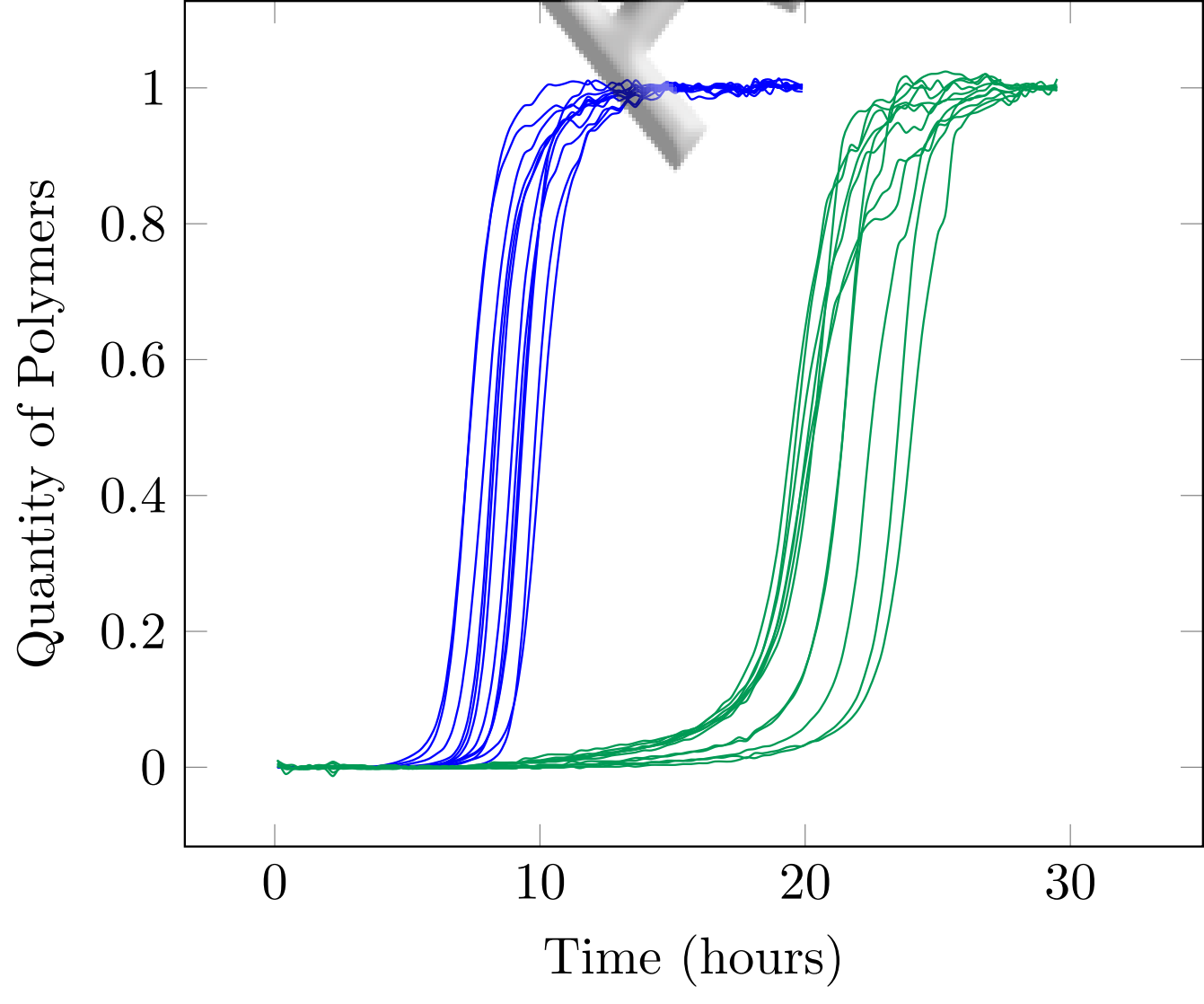
[17] J. Hofrichter, Journal of Molecular Biology **189**, 553 (1986).

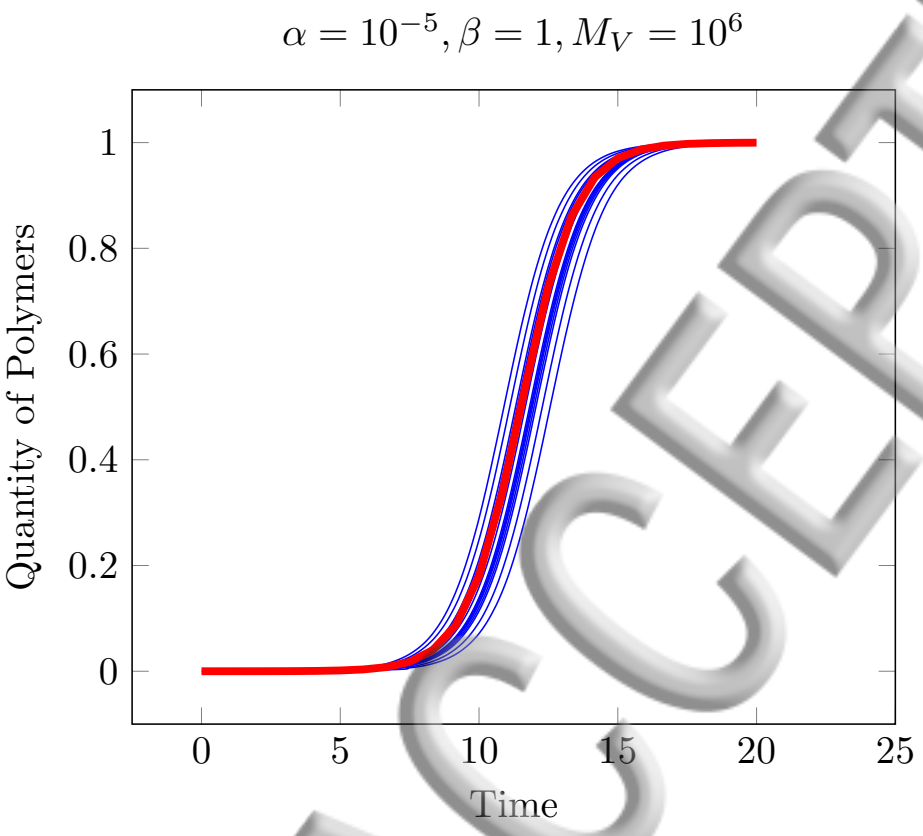
[18] S. Eugène, W.-F. Xue, P. Robert, and M. Doumic, "Supplemental material," (2016).

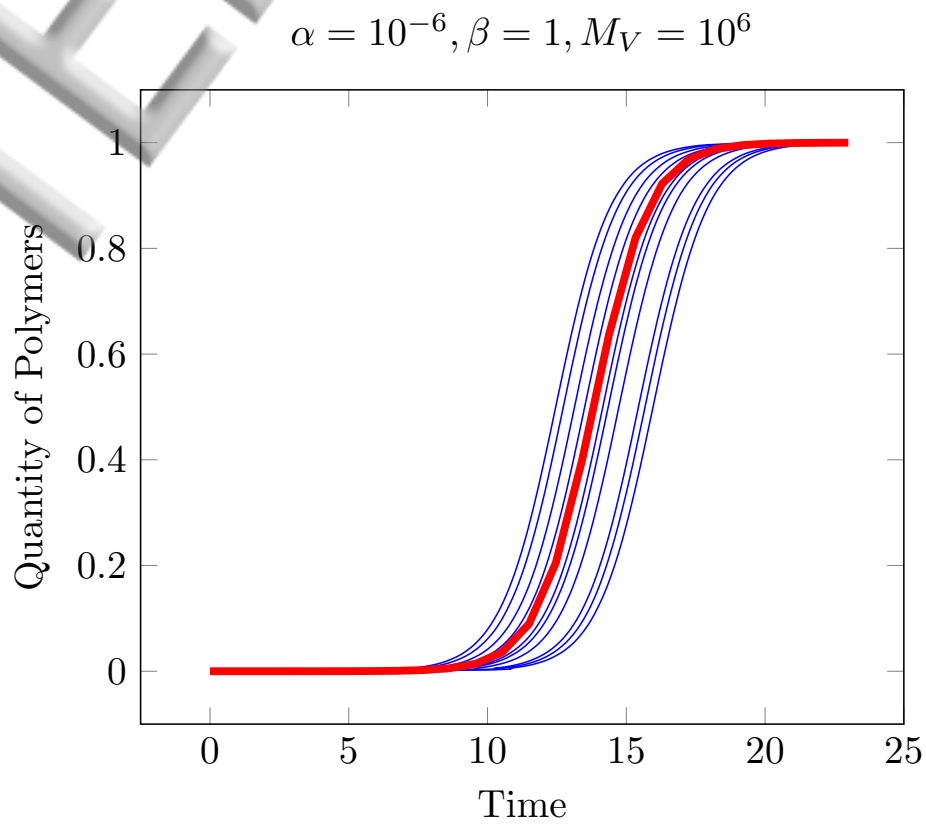
[19] S. Prigent, A. Ballesta, F. Charles, N. Lenuzza, P. Gabriel, L. Tine, H. Rezaei, and M. Doumic, PLoS ONE **7**, e43273 (2012).

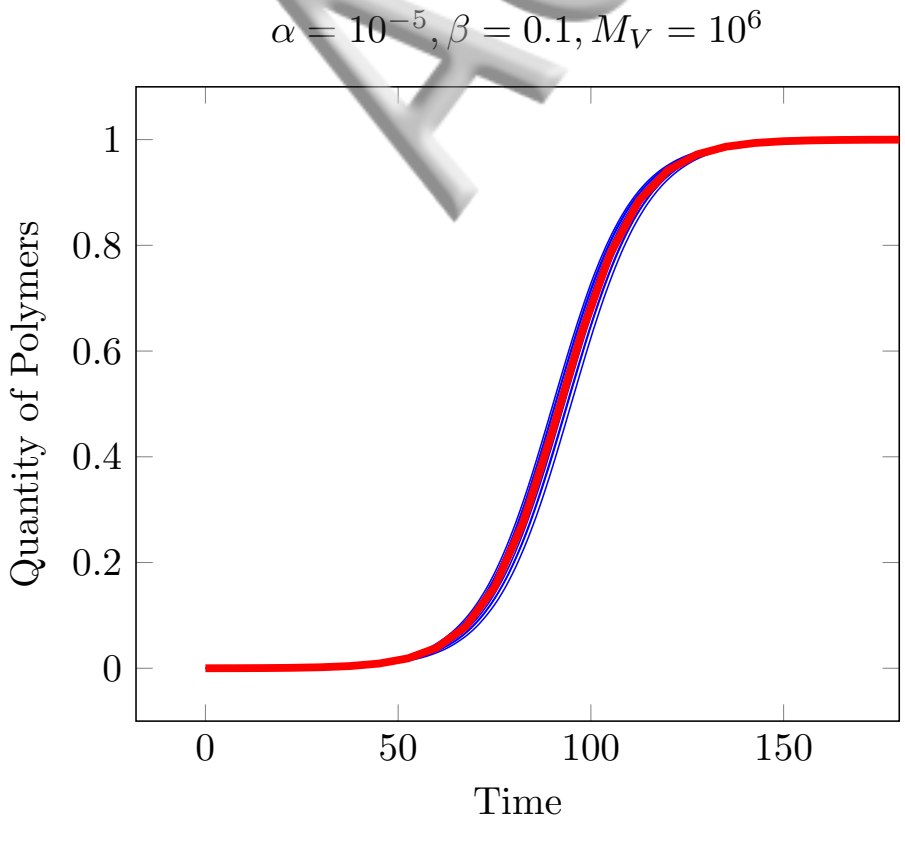
[20] S. N. Ethier and T. G. Kurtz, *Markov processes: Characterization and convergence* (John Wiley & Sons Inc., New York, 1986) pp. x+534.

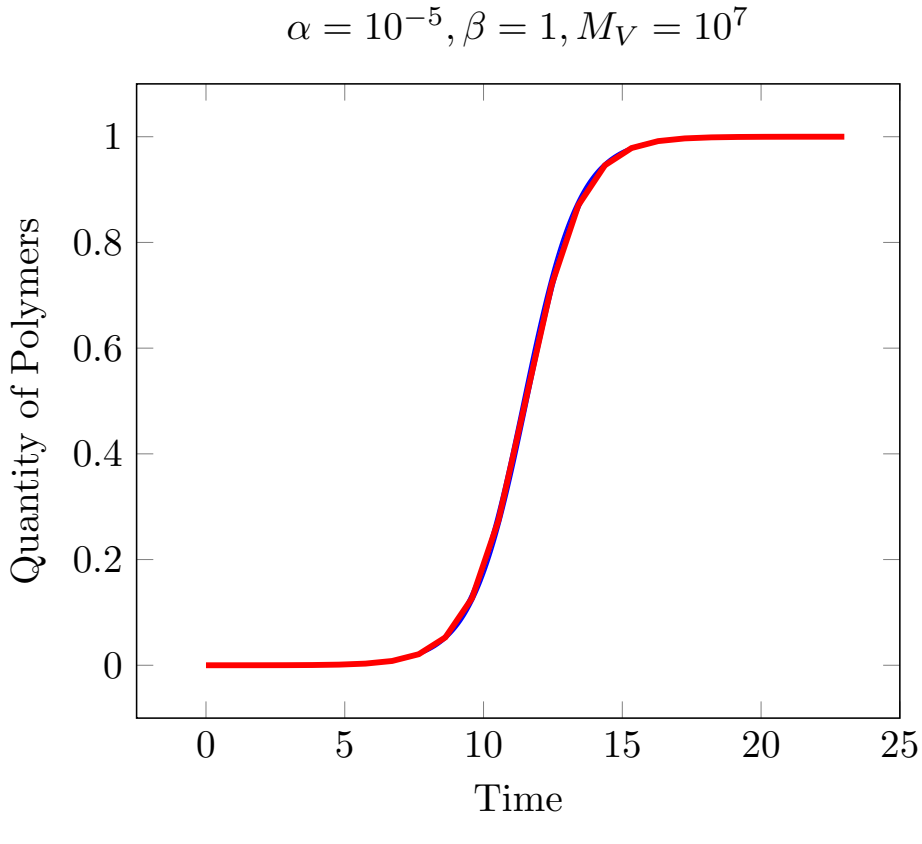

A
B
α
β
Monomer
Monomeric unit in the amyloid state


Quantity of Polymers
1
0.8
0.6
0.4
0.2
0
0
10
20
30
Time (hours)

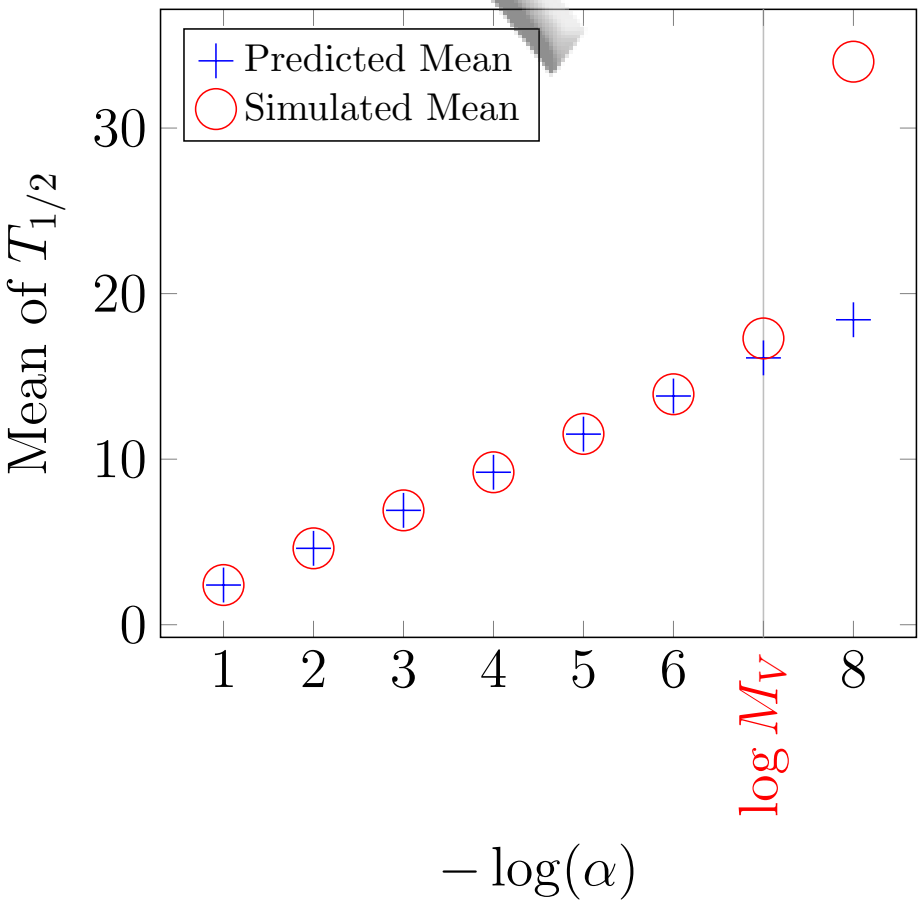

$\alpha = 10^{-5}, \beta = 1, M_V = 10^6$
Quantity of Polymers
Time
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25

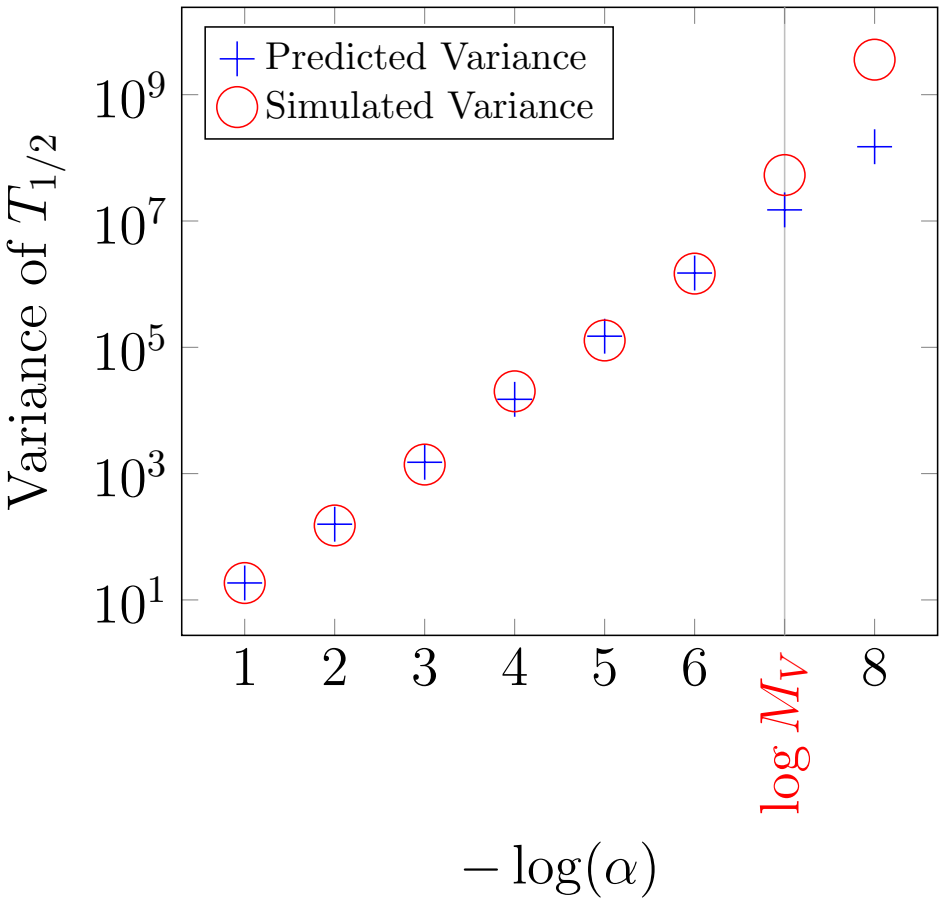

$\alpha = 10^{-6}, \beta = 1, M_V = 10^6$
Quantity of Polymers
Time
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25

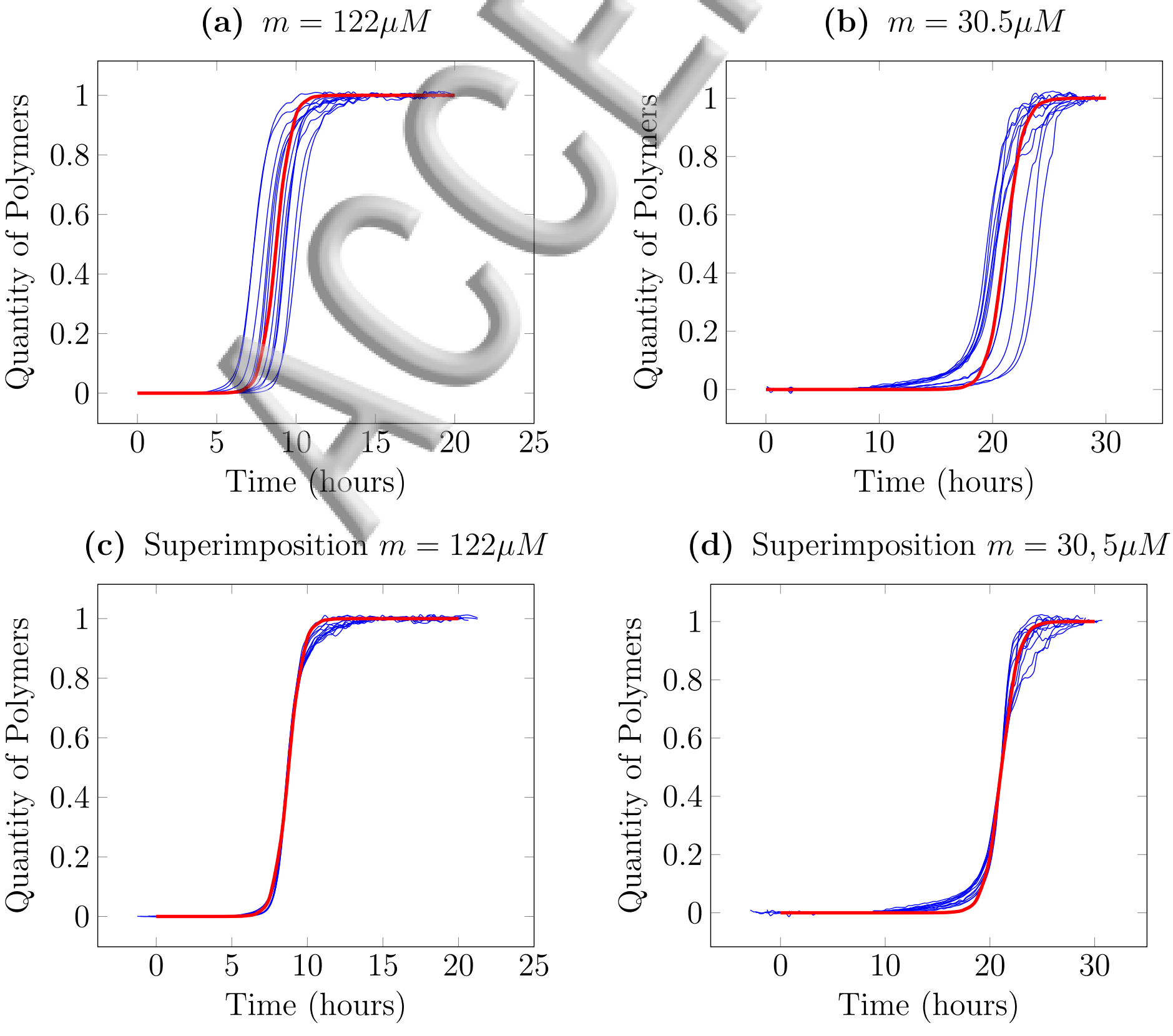

$\alpha = 10^{-5}, \beta = 0.1, M_V = 10^6$
Quantity of Polymers
Time
0
0.2
0.4
0.6
0.8
1
0
50
100
150


$\alpha = 10^{-5}, \beta = 1, M_V = 10^7$
Quantity of Polymers
Time
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25

Regime of Validity of the Model






(a) $m = 122\mu M$
(b) $m = 30.5\mu M$
(c) Superimposition $m = 122\mu M$
(d) Superimposition $m = 30,5\mu M$
Quantity of Polymers
Time (hours)
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25
30