Stochastic nucleation-polymerization model for spontaneous prion protein aggregation

Romain Yvinec [1,*], Samuel Bernard[1], Laurent Pujo-Menjouet[1]
1 Université de Lyon, CNRS UMR 5208 Université Lyon 1 Institut Camille Jordan, 43 Blvd. du 11 novembre 1918 F-69622 Villeurbanne CEDEX, France
∗ **E-mail: yvinec@math.univ-lyon1.fr**

Abstract

We study the aggregation dynamics of the prion protein. We focus on the sporadic appearance of heterogeneous amyloid strain in *in vitro* polymerization experiments. We introduce and analyse a stochastic version of the Lansbury nucleation-polymerization model. This stochastic model is able to reproduce experimental observations where the deterministic model fails. Influence of key parameters is investigated, and we prove that nucleation times can be strongly dependent, weakly dependent or even nearly independent of the initial quantity of monomers relatively to the different values of aggregation kinetic parameters and nucleus size. We show that our discrete and stochastic approach leads to some crucial and new complex behaviour in the nucleation formation process: non-monotonicity of the mean nucleation time with respect to kinetic parameters, weak dependence with respect to initial quantity of monomers, and bimodal nucleation time distribution. These findings may help to understand experimental observations. We finally explain how this stochastic nucleation model can be seen as the building block of a model that would explain how amyloid strain heterogeneity could arise from an homogeneous solution of prion protein.

Author Summary

This paper deals with the study of a stochastic discrete model based on biochemical reactions of nucleation, lengthening and fragmentation. The introduction of such model is motivated by the observed heterogeneity of the dynamics of spontaneous polymerization of prion protein of *in vitro* experiments. This model can be seen as a counterpart of deterministic models widely applied for protein aggregation dynamic such as the Becker-Döring model or other nucleation-polymerization models. Such an approach has never been applied to protein aggregation dynamic. The discrete and stochastic aspects of this model allows us to explicitly evaluate the nucleation time, that is the first time a single nucleus (or aggregate) of a critical size appear. We extensively study the statistics of the random variable defined by this nucleation time and we obtain various analytical approximations that are compared with numerical simulations. We show that this stochastic model helps to identify parameters such as the nucleus size - by analysing the histogram of nucleation time obtained through repeated spontaneous polymerization *in vitro* experiments. Furthermore, we prove that in a specific region of parameters, the mean nucleation time is very weakly dependent on initial quantity of monomers, which would explain some recent *in vitro* experiment observations obtained on spontaneous prion polymerization. We finally indicate how such study can be extended to a stochastic nucleation-dependent polymerization model, and could explain how experimentally observed heterogeneous polymer strains may emerge from an homogeneous pool of monomers.

Introduction

Diseases such as Creutzfeldt-Jacob or Kuru for human, and bovine spongiform encephalopathies (BSE), scrapie or chronic-wasting disease for animals are all spongiform encephalopathies and belong to a

larger class of neurodegenerative disorders [1, 2]. It has generally been accepted that spongiform encephalopathies result from the aggregation of an ubiquitous protein, the prion protein, into amyloids [3–5]. Formation of prion amyloid is believed to originate from a change of conformation of the prion protein [6, 7]. The normal, or non-pathological conformation of this protein is called PrP^c, standing for Common or Cellular Prion Protein. The abnormal, or pathological conformation is called PrP^{c*}. This misfolded protein has a tendency to form aggregates. These aggregates are called PrP^{Sc}, standing for Scrapie Prion Protein. The many different pathways leading to amyloid fibril formation from single proteins (monomers) or pre-formed seed (polymers) are not fully understood and still subject to controversy [8–17].

The neurodegenerative process could be due either to the specific form of aggregates (gain of toxic functions) [18] or to the depletion of native PrP^c monomer (loss of functions) or both phenomena. Protein depletion would indeed be the consequence of polymerization into PrP^{Sc} polymers after a pathological conformation change. PrP^c monomers can take numerous different conformations, which could theoretically produce as many PrP^{Sc} polymers strains. It has been observed [19, 20] that each polymer strain can possess different kinetics under the same experimental conditions: polymerization rates or nucleation time (time required for a first nucleus to be formed) could vary. Moreover, it has even been shown that some strains are infectious for some of the mammalian species and totally harmless for other populations (even within a same species) [8, 21–23].
A better understanding of strain formation is crucial for finding conditions under which "safe" and pathological PrP^{Sc} polymers can be produced and how infectious they could be.
Two questions arise: is it possible to get coexistence of several strains produced from the same homogeneous population of PrP^c monomers? If so, is it possible to reproduce this phenomenon within a computational framework? The answer to the first question is yes [19, 20]. The answer to the second question is then one of the essential key elements to be taken into account in order to get the consistency of our model. We see below that this is possible.

We present here our mathematical model for PrP^{Sc} protein *in vitro* formation based on the nucleation/polymerization model originally introduced by Lansbury *et al.* in [3]. It is the simplest model we can build that accounts for protein conformational change, nucleation barrier and fast amyloid formation through polymerization and fragmentation. The model consists in a set of chemical reactions, summarized in Figure 1. First, PrP^c monomers undergo reversible changes of conformation through a spontaneous misfolding reaction (Figure 1A). Second, only the misfolded proteins PrP^{c*} are then active in the aggregation process and start to aggregate. These small size aggregates also called oligomers undergo reversible aggregation and lengthening through addition/depletion of one monomer, as in the Becker-Döring model, up to a fixed maximal size (Figure 1B) also called nucleus size. Third, once this critical nucleus size has been reached, PrP^{Sc} polymers become stable, and they grow thanks to an irreversible successive addition of single monomer (Figure 1C). This lengthening process is called polymerization. Finally, large polymers can split into two pieces, not necessarily of equal sizes, through a fragmentation process (Figure 1D). For the sake of simplicity, all kinetic rates are assumed to be independent of the size of the aggregates, so that the same parameter value is valid for each kinetic step (see Table 1).

The mathematical formulation of this problem is given by a stochastic discrete model based on a pure-jump Markov process [24]. In such models, each variable is represented by a discrete number and evolves at random discrete time steps when a reaction happens. Propensities of the reactions govern the mean frequency at which reactions occur. This choice of a stochastic discrete model enables us to define nucleation as a discrete event, that is the first time a nucleus of a critical size is formed. Our choice of a relatively simple model is made in order to proceed to a satisfactory mathematical analysis and to understand the effect of stochasticity in this spontaneous polymerization process. Indeed, one of our goals is to look for an accurate characterisation of the nucleation time (in terms of the kinetic

parameters and the total number of proteins) that can be further compared to experimental data, as well as a characterisation of the polymerization process once a first nucleus has been formed.

Consequently, the use of various mathematical simplifications allows us to derive analytic results showing the dependence of the nucleation time with respect to parameters. We show in the next sections that the stochastic nucleation model displays unusual behaviour. Indeed, within a large range of parameters, our investigations show that nucleation time may be roughly independent of the initial quantity of PrP^c monomers. Hence, our work indicates that experimental PrP^{Sc} formation based on data published Alvarez-Martinez *et al.* in [20] can be suitably consistent with our model that includes stochastic terms and discrete variables in the early aggregation steps.

The heterogeneity on prion structure necessary to explain heterogeneity observed in the polymerization dynamics [20] can be induced by our stochastic model. Furthermore, consistent with the these experiments, we show that such a model allows the emergence of different polymer strains from a homogeneous pool of monomers.

1 Results

1.1 Stochastic model

In this section, we introduce our stochastic model of protein polymerization dynamics under the form of a stochastic aggregation-lengthening polymer model. It can be seen as a mixture of a stochastic version of a Becker-Döring model with fixed maximal size and a stochastic model of polymerization-fragmentation.

Description of the chemical reactions Our model consists in a set of chemical reactions involving only PrP^c, PrP^{c*} and PrP^{Sc} prion proteins. They can be summarized in three main stages describe here.

1) First stage. We assume that a PrP^c monomer is able to spontaneously (or sporadically) misfold in a reversible way. The misfolded form, PrP^{c*}, is believed to be very unstable [25], and this process of folding/unfolding is a very fast and frequent event (see Figure 1-A). Moreover, we assume that any misfolded PrP^{c*} monomer is able to actively contribute to polymerization, by adding one misfolded PrP^{c*} monomer to a PrP^{Sc} polymer at each step. The reaction set composed of the reversible folding/unfolding reaction is called the misfolding process.

2) Second stage. The next step deals with the first nucleus formation, under the so called nucleation process. This step consists in only agregation and lengthening reaction of a single misfolded PrP^{c*} monomer. The polymerization reaction rate is denoted k^+, and the backward depolymerization reaction rate k^-. For simplicity in our study, these k^+, k^- reaction rates are supposed to independent of the aggregates size or any other parameters involved in the reaction. This may be less biologically relevant, but the problem would be then very technical and challenging from an analytic point of view. The small aggregates created during this early nucleation process are called oligomers and consist in a number of monomers smaller than a given threshold size N (see Figure 1-B). At this specific size N, oligomers of size larger than N are considered stable, so that the depolymerization reaction becomes impossible. And so, aggregation of any misfolded monomer to this polymer of size larger than N becomes irreversible (see Figure 1-C). This critical oligomer size N at which kinetic steps change is called nucleus size. For the same sake of simplicity, we use a constant-size nucleus size model.

3) Third stage. Dynamics of polymers larger than the nucleus follows then a classic irreversible polymerization-fragmentation model [2], resulting in rapid fibril lengthening (see Figure 1-D). Note that each fibril can split into two parts of different size. And if one of the resulting size is less than the critical nucleus size, this fragmented fibril becomes unstable again and it splits into small misfolded monomer (see Figure

1-D). Again, polymerization rate k_p and fragmentation rate k_b are supposed to be independent of the polymer size for the sake of simplicity.

Variables and Parameters Our model involves a total of 8 kinetics parameters given in Table 1. We distinguish the number M_I of native PrP^c monomers (that are inactive in the aggregation reaction) with the number M_A of misfolded PrP^{c*} monomers (that are active in the aggregation reaction). We suppose that each native PrP^c monomer can fold or unfold with respectively rates γ and γ^*. In some parts of our study below, we may assume that this folding process is fast enough to consider only the equilibrium state and we denote it by $c_0 = \gamma/\gamma^*$. We denote by P_i the number of each polymer of size $i \geq 2$. As explained before, the polymerization rate k^+, and dissociation rate k^- in the nucleation steps are supposed independent of the size $i \leq N$, where N is the critical nucleus size. In some parts of our study below, we rescale the equilibrium rate, and denote it by $q = k^-/k^+$. Finally, once the fibrils are formed and of size larger than N, we denote by k_p the polymerization rate (independent of its size $i \geq N$), and by k_b the PrP^{Sc} polymer splitting rate. The fragmentation kernel is taken as a uniform kernel, that is the size of one of the two polymers resulting from a fragmentation event is picked according to a uniform law among all the possible lower sizes. Finally, we denote by M the total mass of the system (that is the total number of proteins present in the system, in all various forms PrP^c, PrP^{c*} and PrP^{Sc}).

Stochastic formulation of the model We now detail the way we implement our set of chemical reactions in a stochastic model. We model the time evolution of the *discrete* number of each element, and each reaction event is explicitly expressed. Formally, this stochastic model is based on a continuous time Markov chain. The space state of the Markov chain is finite and included in $[\![0; M]\!]^{M+1}$ (note that there is $M+1$ species and each specie cannot exceed M). Each reaction is now interpreted as a possible transition for the Markov chain. Hence, the model is fully defined by the propensity and state-transition of each possible reaction. All the possible transition are summarized in Table 2. Initial condition is always taken as a pure PrP^c monomer state, to reproduce spontaneous polymerization experiments. This model can be simulated according to a Gillespie algorithm [26]. Exact sample paths of the stochastic model are calculated by this computational method, by successively choosing one reaction among all possible according to their propensities. The chosen reaction is executed (which modifies the number of the involved species), the time is updated, and a new reaction is chosen, and so on. We present a typical simulation in figure 2. This illustrates the different steps of the model : namely the misfolding, nucleation and the polymerization-fragmentation steps. An important property of such models is mass conservation, that is the total number of monomers (the free PrP^c and PrP^{c*} ones, and the ones composing PrP^{Sc} polymers) is constant over time. Indeed, since we are dealing with *in vitro* experiment, we assume that there is no source of monomers, and no degradation either. Theoretically, time evolution of the probabilities to be in a particular state are governed by the master equation [24]. However, due to the high dimension of the state space and the complexity of the different possible states in the system, we do not follow the master equation approach to study our model. We use instead an averaging and limit theorem strategy as in [27] to simplify our model, using various assumptions such as time-scale separation, to be detailed in the next paragraph. A simpler linear model, and a deterministic model is then solved analytically, and numerical simulations validate the different approximations performed in the next sections.

1.2 Fast misfolding process results in an average fraction number of active PrP^{c*} monomer

The misfolding process between normal PrP^c and abnormal PrP^{c*} prion monomer conformations is an additional step taken into account in our model, which to the best of our knowledge has never been considered in similar models before. This step, as we have mentioned before, is quite fast since the

folding/unfolding of one monomer is a rapid event in comparison with the other molecular interactions. It introduces therefore an additional time scale to our system, which becomes analytically challenging. It appears then natural and necessary, that for technical reasons, we simplify this new system and reduce the number of our parameters, starting with this folding/unfolding step. Such a simplification is standard in the deterministic ordinary differential equation theory [28]. However, it has been recently used in stochastic equations as well (see [27] for instance).
In our paper, we use a similar method to tackle our problem. That is, we assume that the two monomer folding reaction rates γ and γ^* involved in the conformation change are fast enough to reach almost immediately an equilibrium in comparison with the other reaction time scales. This hypothesis, is given by $(H1)$:

$$(H1)\begin{cases} \gamma \gg 1, \ \ \gamma^* \gg 1, \\ c_0 = \dfrac{\gamma^*}{\gamma} < \infty. \end{cases}$$

Consequently, it we take $(H1)$ into consideration in our nucleation-polymerization model (detailed in the previous section 1.1 and sum up in Table 2), the associated fast subsystem consisting of PrP^c (correctly folded) monomers, whose number is denoted by M_I, and PrP^{c*} (misfolded) monomers, whose number is denoted by M_A, has a unique equilibrium distribution, only depending on the total quantity of monomer $M_F(t) := M_I(t) + M_A(t)$ and is given by the following binomial distribution of parameter $M_F(t)$ and $\frac{\gamma^*}{\gamma+\gamma^*}$:

$$M_I(t) \sim B(M_F(t), \frac{\gamma^*}{\gamma+\gamma^*}),$$
$$M_A(t) = M_F(t) - M_I \sim B(M_F(t), \frac{\gamma}{\gamma+\gamma^*}).$$

The binomial distribution is a consequence of the first-order reaction kinetic assumed in our model for the folding process, and reflects the fact that every monomer has a probability $\frac{\gamma^*}{\gamma+\gamma^*}$ to be correctly folded, and a probability $\frac{\gamma}{\gamma+\gamma^*}$ to be misfolded, independently of the other monomers (as in a toss-coin model). So, if $(H1)$ holds, M_A is a fast switching variable and the quasi steady-state first two moments are given by

$$< M_A > (t) = M_F(t)\frac{\gamma}{\gamma+\gamma^*},$$
$$< M_A(M_A-1) > (t) = M_F(t)(M_F(t)-1)\Big(\frac{\gamma}{\gamma+\gamma^*}\Big)^2,$$

and so, the mean number of misfolded protein and its variance are proportional to the total quantity of monomer and its variance, respectively. We can reduce further the number of parameters in our model with the following. Rescaling the time as being $\tau = \frac{\gamma k^+}{\gamma+\gamma^*}t$, and using the following notations

$$q := \frac{k^-}{k^+},$$
$$c_0 := \frac{\gamma^*}{\gamma},$$
$$q_0 := q(1+c_0),$$
$$K_b := \frac{k_b}{k^+}(1+c_0),$$
$$K_p := \frac{k_p}{k^+}(1+c_0),$$

the resulting stochastic nucleation-polymerization system consists in a *single* monomer species (whose number is denoted by M_F) and polymer species (P_i, $i \geq 2$). Hence, the state-space of the model is now

included into $[\![0; M]\!]^M$ (as in traditional coagulation-fragmentation model, one variable for each size). Kinetic rates are adjusted to take into account only the fraction of monomers that actively participate to the nucleation process. The resulting set of transitions in this model with condition (H_1) are summarized in Table 3, and fully define a new Markov chain model.

1.3 Nucleation lag time analysis reveals three distinct bevahiors

Nucleation lag time corresponds to the time when the very first nucleus (an aggregate of size N, by definition) is formed. The nucleation lag time is defined as follows:

$$T = \inf\{t \geq 0, P_N(t) = 1\}. \tag{1}$$

Let us denote by $< T >$ its mean values, $u_T(t)$ its probability distribution and $S_T(t) = \int_t^\infty u_T(s)ds$ its tail distribution.

The simplifying assumption $(H1)$ in the reduced system (see section 1.2 and Table 3), allows us to deal with only four parameters during the nucleation process, that are namely the initial total number of monomers M, the equilibrium monomer folding rate c_0, the equilibrium polymerization/dissociation reaction rate q involved in the nucleation step, and the nucleus size N. The other parameters (K_p, K_b) are only relevant when at least one nucleus has been formed (see section 1.4).

Case $c_0 = 0$ Let us discuss the nucleation lag time statistics for the reduced stochastic model (defined in Table 3) with $c_0 = 0$, for mathematical convenience (which correspond to an homogeneous pool of misfolded monomer). We are aware that it seems obviously more biologically relevant to consider $c_0 > 0$ in a biological system. This latter case will be analysed further down in this paper.
Let us consider now only the case $c_0 = 0$. In such a case, there is no folding / misfoding exchange between monomers and the model appears to be a conservative Becker-Döring model with constant kinetic coefficient (independent of the aggregates size), and with fixed maximal size, given by N (as we are only interested in the nucleation time here). The statistic of the nucleation lag time as a function of the aggregation / disaggregation kinetic rate q for the stochastic version of the Becker-Döring model has been studied in a companion paper [29]. In that paper, it has been shown that:
1) In the unfavourable case for nucleation formation, when $q \gg M$, a pre-equilibrium assumption (that is all oligomer of size less than N are in equilibrium states) becomes valid and the lag time distribution is close to an exponential distribution. The parameter of the exponential is approximated by

$$\frac{2q^{N-2}}{M^N}. \tag{2}$$

In general, however, the parameter of the exponential distribution is given by the second order asymptotic moment value

$$< M_F P_{n-1} > (t \to \infty) \tag{3}$$

of a Becker-Döring model with finite maximal size equal to $N - 1$. Such asymptotic moment is given by the previous expression (2) only when correlation between variables can be neglected, and when the mean value is closed to the deterministic model due to standard mean-field hypotheses. The exact computation of such second order moment value seems to be an unreachable problem, but it can however be sorted out with numerical simulations (see numerical results and discussion in [29]) or moment closure techniques [30]. The exponential function for the nucleation time distribution arises in the limit of fast aggregation/disaggregation rate $q \gg M$ because the reaction $M_F + P_{N-1} \to P_N$ is occurring at a time-scale much slower than any other reactions in the system. As a single firing occurrence of such reaction is needed for the nucleation event to appear, the exponential function is then just a consequence of the

Markovian property of the stochastic model.
2) The mean nucleation time is a non-monotonic function of the aggregation/disaggregation rate q, and $< T > \to \infty$ as $q \to 0$. Hence, if we start with a small value of q, then increasing this rate would lead to a *shorter* mean nucleation time. This unexpected effect arises from a redistribution of trajectory weights such that upon increasing the rate q, paths that take a shorter time to complete a cluster become more likely. This effect is also the consequence of single-monomer aggregation hypothesis. For small rate q, trajectories for which only few (or none) free monomers remains before a nucleus is formed leads to a longer nucleation process, as some aggregates must lengthen (at speed roughly of $1/q$) to free some monomers so that other aggregates can grow again. For larger q, the quantity of free monomers is high enough to allow the aggregates to grow up to the nucleus size.

Analytical result: in the favourable aggregation limit, as $M \gg q$, it is possible to obtain analytic approximation, which we now present here. This completes then the picture of the nucleation time analysis, and allow us to deduce the behaviour of the nucleation time as a function of the total initial quantity of monomers M. As written in the introduction, such relationship between nucleation time and initial quantity of monomers is important as it can be obtained experimentally.

As the initial number of monomers is large, the early dynamics of the stochastic model behaves as a deterministic model, with a standard mean-field hypothesis. It is known [31, 32] that under favourable aggregation limit assumption, a deterministic Becker-Döring model exhibits the following successive periods:
- firstly, the model behaves as irreversible aggregation process during a time-scale of order $(e/M)log(1/q)$, until monomer concentration becomes small;
- secondly, when the monomer concentration is of order q, there is a metastable period in which each concentration species of size $i \geq 1$ are nearly constant, equal to (by definition) p_i^* (see Figure 5). Those values are distinct from steady-state values;
- thirdly, at time scale of order $1/q$ (which is the time scale of aggregates lengthening), larger aggregates are created within a process akin to diffusion in the size i-space (slow redistribution of aggregates size);
- Finally, every concentration species reach their steady-state values within a time scale of order $1/q^2$.

For an accurate approximation of the nucleation time (*i.e.* the time to reach the first nucleus), we need to understand in which of these periods a first nucleus appears. This mostly depends on the critical size N of the nucleus as follows.
Here is how we do proceed: in order to compute this approximation, we compute the metastable deterministic values p_i^* for our model (see subsection 3.1 and Table 4). Indeed, if the stochastic model follows closely the deterministic model, the nucleus number $P_N(t)$ (which is an integer in the stochastic model) will be close to p_N^* during the metastable period. Hence, for small enough N such that $p_N^* \gg 1$, we expect $P_N(t)$ to reach one in the pure-aggregation period (that is before the metastable period). In the opposite case, for large enough N such that $p_N^* \ll 1$, we expect $P_N(t)$ to reach one after the metastable period. We now distinguish between the two cases and provide an analytical approximation for both cases.

a) In the small nucleus size scenario, $p_N^* \gg 1$, a first nucleus appears while there is still a large number of monomers. With a crude approximation, we can treat the number of monomers as a constant number, equal to its initial value M. In this constant monomer formulation (see subsection 3.2), the stochastic model becomes linear (that is all the reaction propensities are linear functions of the species numbers). Such linear model is analytically solvable (see subsection 3.2), and we can deduce the nucleation time distribution. Hence, we find that in the favourable case $M \gg q$ and small nucleus size scenario $p_N^* \gg 1$, the nucleation time distribution u_T in the stochastic model is approximated by a Weibull distribution and the mean nucleation time is given by

$$< T > \sim \frac{(2(N-1)!)^{1/(N-1)}}{M^{N/(N-1)}}. \tag{4}$$

b) In the large nucleus size scenario, $p_N^* \ll 1$, the relevant time scale is given by the metastable period and is of order $1/q$. Therefore, we can neglect the initial pure-aggregation phase, and use the metastable values p_i^* as initial values for a linear model where the monomer number is constant and equal to p_1^* (see Figure 5). Again, this linear model can be solved analytically and used as an approximation of the stochastic model. Hence, in the favourable case $M \gg q$ and large nucleus size scenario $p_N^* \ll 1$, the nucleation time depends on the value of p_1^*, given by (see subsection 3.1)

$$p_1^* = q\frac{p_2^* + \sum_{i=2}^{N-1} p_i^*}{\sum_{i=2}^{N-1} p_i^*}, \tag{5}$$

where all p_i^*, $i \geq 2$, are linearly proportional to M (see subsection 3.1). As a consequence, p_1^* is independent of the initial number monomers M and is of order q. Thus the nucleation lag time depends on M only through the initial condition p_i^*, $i \geq 2$, and is found to be (see numerical results in the next paragraph) almost independent of M for N larger than 15. Finally, note that there is a small probability that a nucleus is formed before the metastable period, which we neglect here (see next paragraph).

Hence, in the favourable aggregation limit, as $M \gg q$, we have two different analytical approximations, according to the nucleus size N. We now turn to the numerical study (using Gillespie's algorithm), in order to validate our analytical approximations.

$< T >$ ***versus*** M **and numerical result** Numerical results (see methods in subsection 3.3) confirm the analysis performed above, for both mean nucleation time and nucleation time distribution. Let us summarize here the different approximation we have derived (see also Table 4).

In log scale, the mean nucleation time $< T >$ as a function of the number of initial monomers M has either two or three main behaviours (depending on parameter values) which we now describe (Figure 3, 4).

1) In the unfavourable case, that is for small $M \ll q$, the mean lag time is given approximately by

$$< T > \sim \frac{2q^{N-2}}{M^N}, \tag{6}$$

and the exponential approximation for the nucleation time distribution u_T is valid.

2) For intermediate M and large N, $p_N^* \ll 1$ and the mean lag time is roughly independent of M. In such case the linear metastable approximation is valid. Nucleation time distributions are bimodal in such region, due to the following dichotomy : the first peak is formed by trajectories for which a nucleus appears before the metastable period (during the pure-aggregation period), and the second peak by the trajectories for which a nucleus appears during the metastable period. As the two periods occur at different time scale (the pure-aggregation time scale is $(e/M)\log(1/q)$, the metastable time scale is $1/q$), this leads to two distinct peaks in the nucleation time distribution. The linear metastable approximation captures the second peak of such distribution (Figure 5). And the larger N is, the smaller $p_N^* \ll 1$ is and the weight of the second peak is then more important (it is much likely that the nucleus is formed after the metastable period). The bimodality of the nucleation time distribution is the consequence of the two distinct time scales, and the fact the nucleation event can occur in both time scales in the case $p_N^* \ll 1$. This is not possible in an other parameter region.

3) Finally, for larger M, $p_N^* \gg 1$ and the mean lag time follows approximately

$$< T > \sim \frac{(2(N-1)!)^{1/(N-1)}}{M^{N/(N-1)}}, \tag{7}$$

and the Weibull approximation (given by the constant monomer scenario) for the nucleation time distribution u_T is valid.

Extension for $c_0 > 0$ The case with $c_0 > 0$ (that includes a reversible misfolding process, taken at equilibrium) does not add any analytical challenge, but only modifies criteria and analytic approximations as follows.

- when $q_0 \gg M$, the lag time distribution is close to an exponential distribution of mean parameter

$$< T > \sim \frac{2q^{N-2}}{(1+c0)^N M^N}, \tag{8}$$

- when $M \gg q_0$, a similar dichotomy as in paragraph 1.3 is still valid. Hence for small N, the distribution is approximated by a Weibull distribution and the mean nucleation time is given by

$$< T > \sim \frac{(2(N-1)!)^{1/(N-1)}}{((1+c_0)M)^{N/(N-1)}}. \tag{9}$$

  For larger N, the same linear approximation is valid, with the values of p_i^* changing (see section 3.1), and

$$p_1^* = q_0 \frac{p_2^* + \sum_{i=2}^{N-1} p_i^*}{\sum_{i=2}^{N-1} p_i^*}. \tag{10}$$

These calculus end up our nucleation time analysis. The different behaviors of the model for the nucleation time with parameter criteria are sum up in Table 4. The importance of this deep study for experimental analysis of nucleation time is discussed in the section 2.

1.4 Nucleation formation study suggests two different regimes for polymerization dynamics

We now focus our analysis on the full stochastic discrete model of spontaneous nucleation-polymerization, in order to understand the polymerization dynamics, after the first nucleus has appeared. We still restrict ourselves to the reduced version described in subsection 1.2 and summarize it in Table 3, with instantaneous equilibrium between normal PrP^c and misfolded PrP^{c*} monomers.

In most of nucleation-polymerization-fragmentation models, once some polymers are formed, there exist two ways to create more polymers: either by fragmentation of existing polymers (sometimes referred as secondary nucleation) or by addition of new polymers through spontaneous nucleation.

Our previous analysis of the nucleation lag time (subsection 1.3) suggests two different limits where only one of the two ways to create more polymers is dominant over the other one. Indeed, in the previous section we successively give the cases where $q \gg M$ and $M \gg q$. In the first unfavourable aggregation case, $q \gg M$, appearance of nucleus through spontaneous nucleation is unlikely, and once a first nucleus appears, the dynamic is governed by the fragmentation process leading to an increase of the number of polymers and hence accelerates the polymerization production process. In the opposite favourable aggregation case, $M \gg q$, the appearance of new nucleus is rather governed by spontaneous nucleation, as once a single nucleus is formed, many others are created successively.

Again, for simplicity, as a first step, we take $c_0 = 0$, and then, similar results hold for $c_0 > 0$ due to the pre-equilibrium hypothesis.

1.4.1 Hybrid polymerization-fragmentation approximation for the unfavourable case $q \gg M$

In such limits $q \gg M$, new nucleus are unlikely to appear compared to the fragmentation process. That is, when one nucleus has been formed, its polymerization dynamics becomes faster than the nucleation process of prion proteins and prevents these latter from creating a new nucleus. In this particular case,

almost all monomeric proteins contribute then to the polymer lengthening. To describe this process, we use here a treatment of the stochastic system that takes stochasticity in the fragmentation part into account. This is based on biological assumption (fast growing polymers) where the possibility of nucleation is ignored. We found that the total mass of polymers, $Z(t) = \sum_{i \geq N} iP_i(t)$, can be approximated by (see subsection 3.4) $\hat{z}^n$, where

$$\hat{z}^n = z(t) + \frac{1}{\sqrt{n}}\xi(t), \tag{11}$$

with

$$\left\{ \begin{array}{lcl} \dot{m}(t) & = & -K_p m(t)p(t), \\ \dot{z}(t) & = & +K_p m(t)p(t), \\ \dot{p}(t) & = & +K_b z(t), \end{array} \right. \tag{12}$$

and

$$\left\{ \begin{array}{lcl} \xi(t) & = & -K_p \displaystyle\int_0^t p(s)\xi(s)ds + K_p \int_0^t m(s)\rho(s)ds, \\ \rho(t) & = & K_b \displaystyle\int_0^t \xi(s)ds + W(\int_0^t K_b z(s)ds) \end{array} \right. \tag{13}$$

where W is a standard Wiener process, and n a scaling parameter of the polymerization and fragmentation rate (see subsection 3.4). Thus, the total mass of polymers $Z(t) = \sum_{i \geq N} iP_i(t)$ can be approximated by a Gaussian process $\hat{z}^n(t)$ of mean value $z(t)$ and variance

$$\left\{ \begin{array}{lcl} < \xi^2(t) > & = & \displaystyle\int_0^t K_b z(s) A(s,t) ds, \\ A(s,t) & = & \displaystyle\frac{b^2}{a^2 + 4bc}\Big(e^{\frac{a-\sqrt{a^2+4bc}}{2}} - e^{\frac{a+\sqrt{a^2+4bc}}{2}}\Big)^2, \\ a & = & \displaystyle -K_p \int_s^t p(u)du, \\ b & = & \displaystyle K_p \int_s^t m(u)du, \\ c & = & K_b(t-s). \end{array} \right. \tag{14}$$

Those equations arise from a central limit theorem in the limit of large population of proteins and fast polymerization. The first set of equation (12), which gives the mean behaviour, is the classical deterministic model of polymerization-fragmentation in the case of constant kinetic parameters and uniform fragmentation kernel (written with the use of aggregate variable, see [33]). This deterministic model describes the time evolution of the number of monomers ($m(t)$), the mass of polymers ($z(t)$) and the number of polymers ($p(t)$). The mass of polymers grows proportionally to the number of encounters of monomers and polymers, given by $m(t)p(t)$. The number of polymers grows by fragmentation of existing polymers. The fragmentation occurs at a rate proportional to their mass.

In our approximation, the fluctuations around the mean behaviour, given by the second set of equations (13), are governed by the fragmentation process, the variance of mass of polymers being proportional to the fragmentation rate (equation (14)).

It is important to note that such analytic expression is of practical importance thanks to the work provided by Knowles *et al.* in [34]. The authors were indeed able to give an accurate formulation of the solutions of the deterministic equation (12) using a fixed-point iteration method. They obtained the following expressions:

$$\left\{ \begin{array}{lcl} z(t) & = & z(0) + m(0) - m(0)\exp\Big(-\frac{K - pc_1}{\kappa}(e^{\kappa t} - 1) + \frac{K_p c_2}{\kappa}(e^{-\kappa t} - 1)\Big), \\ p(t) & = & p(0) + c_1(e^{\kappa t} - 1) + c_2(e^{-\kappa t} - 1). \end{array} \right. \tag{15}$$

where

$$\begin{cases} c_1 &= \frac{1}{2}(p(0) + \frac{K_b}{\kappa} z(0)), \\ c_2 &= \frac{1}{2}(p(0) - \frac{K_b}{\kappa} z(0)), \\ \kappa &= \sqrt{K_b K_p M} \end{cases} \tag{16}$$

Consequently, this derivation allows us to solve equation (13) and to get an accurate expression of $\hat{z}^n$, which approximates the total mass of polymers $Z(t)$ in the full discrete stochastic model (Figure 6), in the limit $q \gg M$. This study may be used to get an expression of the lag time required for the total mass of polymers to reach a given fraction $h \in (0, 1)$ of the total mass in the system (such quantity is used as a measure of the nucleation time in a continuous model [20]), with

$$T_h := \inf\{t \geq 0 : Z(t) \geq hM\} \approx \hat{T}_h := \inf\{t \geq 0 : \hat{z^n}(t) \geq hM\}. \tag{17}$$

The statistic of the latter quantity is easily derived by

$$\mathbb{P}\big\{\hat{T}_h \geq t\big\} = \mathbb{P}\big\{\xi(s) < \sqrt{n}(hM - z(s)), 0 \leq s < t\big\}. \tag{18}$$

Although we do not pursue on this here, note that such quantity may also be used to quantity the characteristic time of the polymerization process, or the polymerization speed, in a discrete stochastic model.

1.4.2 Nucleation dominant limits for the case $M \gg q$

In the opposite case, $M \gg q$, we see that a large quantity of oligomers is created during the nucleation period, which may eventually lead to many nucleus and polymers after a slow metastable phase of length $1/q$. Once this phase ends, oligomers continuously lengthen into polymers adding more and more polymers. Then, the existing polymers do not polymerize as fast as new nucleus appears, because free monomers are not present in large quantity (of order $q \ll M$). Thus, the overall polymerization process is dictated by oligomer disappearance (and creation of new nucleus) rather than polymer growth and fragmentation, and occurs at a speed of order $1/q$.

We illustrate the deterministic nucleation dominant approximation with several realizations of simulations (Figure 7). As a consequence, polymerization kinetic parameters K_p and K_b have little influence in such case as long as they are sufficiently large (so that they do not limit the polymerization process). In other words, polymer mass time evolution follows a deterministic curve, shifted from one another due to the variability in the nucleation time.

2 Discussion

In this paper we have analyzed the statistics of the nucleation time in a stochastic Becker-Döring nucleation model. We have also extended our analysis to a nucleation-dependent polymerization model.

Literature review on lag time studies. A few stochastic model have been applied to protein aggregation kinetics in general, and prion amyloid formation in particular. One major focus of these studies was the distribution shape of the lag time, defined in all quoted work as the time for the polymer mass to reach a given critical level [35–38]. In [36], the authors used a simple autocatalyic conversion kinetic model to obtain the distribution of incubation time. Under the assumption that the constant rate involved is a log-normally distributed stochastic variable, the incubation time is then also shown to be log normal distributed. In [35, 37], the authors derived the distribution shape of lag time using assumptions

on probabilities of nucleus formation event. Hofirchter [37] exhibited a delay exponential distribution, while Szabo [35] obtained a β-distribution, useful to experimentally deduce the rate of a single nucleation formation. In [38] the authors derived a phenomenological model to get the mean time needed to obtain a certain amount of polymer, given one initial seed and given assumptions of distribution of aggregation and splitting times. This expression allowed them to discuss the influence of initial dose or other parameters on the incubation time. Using a purely stochastic sequential aggregation of monomers and dimers model was their strategy to obtain different lag time distributions as a γ-distribution, a β-distribution or a convolution of both.

A stochastic discrete aggregation kinetic model. In our approach, we do not use a phenomenological model, where nuclei are supposed to appear at a given rate, but a purely stochastic aggregation kinetic model, where nucleus formation is the result of successive addition and disassociation of single misfolded monomer. This discrete stochastic model allows us to define a nucleation time to be the time needed to form the first polymer of a given critical size. Up to our knowledge, such definition, although relatively simple, had never been considered for describing protein nucleation kinetics before. Previous initial analysis of this discrete stochastic model were performed in [29,39,40], but the specific dependence of the statistics of the nucleation time with respect to the total quantity of monomers was not considered.

Stochasticity helps identifying parameters (nucleus size). Our analysis has shown that a stochastic Becker-Döring nucleation model leads to three different regimes for the nucleation time, according to parameter values. We have reduced our analysis to three main parameters: the total number of protein M, the normalized dissociation kinetic rate q, and the nucleus size N. In the extreme favourable case where proteins are initially present in a large quantity, $M \gg q$, the mean nucleation time is roughly inversely proportional to the total quantity of monomers, and the nucleation time distribution follows a Weibull distribution. In the extreme unfavourable case, $q \gg M$, the mean nucleation time decreases as $1/M^N$ as the total quantity of monomers M increases, and the nucleation time distribution follows an exponential distribution. These two regimes are well known in the literature, within a deterministic modelling framework [41, 42]. They have implications in terms of identifiability of parameters. In particular, the nucleus size can be identified from experimental data only in the regime $q \gg M$, with a plot of the mean nucleation time versus the total quantity of monomers.
However, one difficulty that remains is to know whether the unfavourable assumption is valid or not. Our analysis shows that experimental nucleation time distribution may help to know in which parameter regime spontaneous polymerization experiments were performed. Indeed, probability distribution are very different (Weibull and exponential) in the two extreme favourable and unfavourable regimes. Hence, a statistical treatment of repeated spontaneous polymerization experiments can extract experimental nucleation time distribution and help to discriminate between both extreme parameter region. Moreover, in the favourable case $M \gg q$, the nucleation time distribution may help identifying the nucleus size as well (while the mean nucleation time cannot help in such case), as the Weibull distribution initially increases at speed t^{N-1}. To be helpful, this requires a high number of repeated experiments to accurately derive an estimate of the nucleation time distribution.

The stochasticity brings new behaviour (and explains experimental data): independence of nucleation time with respect to initial number of monomers. Apart from these two well-known regimes, the discrete stochastic Becker-Döring model displays an additional feature for the nucleation time statistics. In the favourable case, $M \gg q$, and large nucleus size N, the mean nucleation time is very weakly dependent on the total mass of monomers, and the probability distribution is bimodal (this is related to a metastability phenomena, see section 1.3). Up to our knowledge, this behaviour has

not been reported before in a computational framework, and is specific to a stochastic treatment of the Becker-Döring nucleation model.
In vitro polymerization experiments of prion protein give some interesting insights of what could be the different mechanisms involved in the nucleation process. In particular, the observed dynamics of prion polymerization can be compared to the result of mathematical modelling, and different biological hypotheses can be tested. One of the challenges of the prion polymerization experiments resides in the low sensitivity to the dynamical properties of the polymerization on initial concentration of PrP^c protein [12, 42–44]. Our findings may explain recent experimental observations for spontaneous prion polymerization experiments [20], where the mean nucleation time decreases as $M^{-0.5}$ or lower as the initial quantity of monomers M increases. According to our mathematical modelling approach, this suggests a favourable case scenario, $M \gg q$ and large nucleus size $N \geq 15$. One way to test this hypothesis would be to estimate the nucleation time distribution from the experimental data in [20]. To be consistent with our prediction, the estimated distribution should be bimodal. For the reader convenience, and with permission of [20], we have reproduced nucleation time statistics extracted from the experimental data in [20], in Figure 8. The very few numbers of available experiments unfortunately prevents us to derive a serious conclusion from it, but may give an important future direction of research. Another important direction of research includes a generalization of this work for size-dependent kinetic rate, and more general coagulation-fragmentation process.

Extension to a nucleation-dependent polymerization model: the nucleation source term. The extension of our analysis to a nucleation-dependent polymerization model also suggests two different behaviours for such model: one where the fragmentation process dominates the production of new polymers ($q \gg M$), for which a Gaussian approximation is valid; and one where the nucleation process dominates the production of new polymers ($M \gg q$), and the polymerization kinetic parameters K_p and K_b have little effect. Our findings may help to understand/justify the addition of a nucleation source term in a deterministic polymerization-fragmentation model, usually as M^N [16, 33, 34, 45]. However, according to our study, this term is only justified in the pre-equilibrium hypothesis corresponding to the unfavourable case $q \gg M$. Hence, another important consequence of our work (at least from a theoretical perspective) is to justify nucleation source term in deterministic model.

The stochasticity allows heterogeneity of polymer structures to emerge from an homogeneous pool of monomers. Finally, it has been argued [20] that different monomer conformations may lead through the nucleation process to different polymer structures, and thus different polymerization kinetic, as different polymer strains may have very distinct physical and chemical properties for the aggregation process. In spontaneous polymerization experiments, competition for the first nucleus may govern the production of a particular polymer structure rather others. When this first nucleus is created, it gives birth to a fast fibril lengthening process and acts like a steam roller giving theoretically little chances for other strains to get enough time to form their own first nucleus. This may be the reason why coexistence of two different strains has hardly been observed in a same *in vitro* experiment. However, as mentioned in introduction, this phenomenon is not impossible. And such a coexistence can be observed and so should be able to be simulated in a stochastic model. From our point of view, the simplest and best approach is to consider a stochastic model of the nucleation process, where a nucleation event is described by a creation of a given critical size aggregate. The analysis of our stochastic model of nucleation-dependent polymerization is thus the first step to include heterogeneity of polymer structures in a nucleation-dependent polymerization model. In a future model, heterogeneity of polymer structures will then emerge from an homogeneous pool of protein monomers through the competition of different nucleation process, corresponding to different misfolded conformation. The statistic of the nucleation time and the polymerization speed is then crucial to study such a competition model.

3 Materials and Methods

3.1 Computation of the metastable equilibrium values p_i^*

A deterministic version of the Becker-Döring model described in Table 3 can be represented as a set of ordinary differential equation

$$\left\{\begin{array}{lcl} \dot{p}_1(t) &=& -\dfrac{1}{1+c_0}p_1^2 - p_1\displaystyle\sum_{i=2}^{N-1} p_i + 2q_0p_2 + q_0\sum_{i=3}^{N} p_i, \\ \dot{p}_2(t) &=& -p_1p_2 + \dfrac{1}{2(1+c_0)}p_1^2 - q_0p_2 + q_0p_3, \\ \dot{p}_i(t) &=& -p_1p_i + p_1p_{i-1} - q_0p_i + q_0p_{i+1}, \quad 3 \le i \le N-1, \\ \dot{p}_N(t) &=& p_1p_{N-1} - q_0p_N. \end{array}\right. \tag{19}$$

Letting $q = 0$ and $c_0 = 0$, and using $\tau = \int_0^t p_1(s)ds$, the system (19) becomes

$$\left\{\begin{array}{lcl} \dot{p}_1(\tau) &=& -\displaystyle\sum_{i=1}^{N-1} p_i, \\ \dot{p}_2(\tau) &=& -p_2 + \dfrac{1}{2}p_1, \\ \dot{p}_i(\tau) &=& -p_i + p_{i-1}, \quad 3 \le i \le N-1, \\ \dot{p}_N(\tau) &=& p_{N-1}. \end{array}\right. \tag{20}$$

Upon taking Laplace transform, $z_i(s) = \int_0^\infty e^{-s\tau}p_i(\tau)d\tau$, letting N large and using the mass conservation property, we obtain the exact formula

$$\left\{\begin{array}{lcl} z_1(s) &=& \dfrac{2Ms}{s^2+(1+s)^2}, \\ z_i(s) &=& \dfrac{Ms}{(s^2+(1+s)^2)(1+s)^{i-1}} \quad 2 \le i, \end{array}\right. \tag{21}$$

Taking Laplace inverse transform, we have

$$p_1(\tau) = Me^{-\tau/2}\Big(\cos(\tau/2) - \sin(\tau/2)\Big), \tag{22}$$

which goes to 0 as $\tau \to \pi/2$. The exact expression of $p_1(t)$ in the original time scale can now be obtained (at least, numerically) by the inversion of the nonlinear transformation that defines τ. We can proceed similarly for each p_i to obtain an expression for the lag time in the irreversible aggregation period. Also we can use the inverse Laplace transform of (21) and letting $\tau \to \pi/2$ to obtain asymptotic values p_i^*, $i \ge 2$, during the irreversible aggregation period. We give in table 5 the values of p_i^*/M for $i = 2..15$. If $p_N^* = p_N(\tau \to \pi/2) \gg 1$, then a sufficient quantity of nucleus will be reached during the irreversible aggregation period. For instance, for $N = 10$, $p_N^* > 1$ for $M > 6.074110^4$, while for $N = 15$, $p_N^* > 1$ for $M > 1.514910^9$. The value of p_1^* is of order q, and an approximation of the value of p_1^* can be obtained by using the first equation of (19), which gives equation (5).

Finally, remark that with $c_0 > 0$, a similar procedure leads to

$$\left\{\begin{array}{lcl} z_1(s) &=& \dfrac{2M(1+c_0)s}{(1+2c_0)s^2+(1+s)^2}, \\ z_i(s) &=& \dfrac{Ms}{((1+2c_0)s^2+(1+s)^2)(1+s)^{i-1}} \quad 2 \le i, \end{array}\right.$$

3.2 The linear model is analytically solvable

We analyse in this part a modification of the stochastic model described in table 3 where we consider the quantity of free monomer M_F as a constant over time, that is $M_F(t) \equiv M$ for all time $t > 0$. We restrict for simplicity to the case $c_0 = 0$, while results for $c_0 > 0$ follows by a similar method.

The main advantage of the constant monomer formulation is to be analytically solvable (within our specific choice of parameters, independent of cluster size). As M_F is constant, the propensities of the reaction becomes linear, and it is known in this case that the time-dependent probabilities to have a given number of aggregate of size i are given by a Poisson distribution [46]. Such distribution is characterized by a single parameter, its mean for instance. Again, the model being linear, the mean number of aggregate of size i, at time t, is given by the solution of a deterministic ordinary differential equation which can be rewritten as

$$\begin{aligned} \frac{d\mathbf{p}}{dt} &= \mathbf{A}\mathbf{p} + \mathbf{B}, \\ \frac{dp_N}{dt} &= M p_{N-1}, \end{aligned} \tag{23}$$

where

$$\mathbf{p} = \begin{pmatrix} p_2 \\ p_3 \\ \vdots \\ p_{N-1} \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} -q-M & q & & & \\ M & -q-M & q & & \\ & \ddots & \ddots & \ddots & \\ & & M & -q-M & q \\ & & & M & -q-M \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} M^2/2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{24}$$

We give now for the sake of completeness the associate exact results.

General solution of the linear model We give here the general solution of the linear model, introduced in equation (23)-(24). We note that the size of the matrix $\mathbf{A}$ is $N-2$. A general form for $p_i(t)$, $2 \leq i \leq N-1$ is given by

$$p_i(t) = \sum_{k=1}^{N-2} \alpha_k e^{\lambda_k t} V_{i-1}^{(k)} - (\mathbf{A}^{-1}\mathbf{B})_{i-2},$$

where $\lambda_k = -(M+q) + 2\sqrt{Mq}\cos(\frac{k\pi}{N-2})$ are the eigenvalues of $\mathbf{A}$, $V^{(k)}$ the associated eigenvector, ($V_i^{(k)}$ denotes its i^{th} component), and α_k are constant given by the initial condition. Exact formula for $V^{(k)}$ are given by [47]

$$V_i^{(k)} = \sqrt{\frac{M}{q}} \sin(\frac{ki\pi}{N-1}),$$

and the vector $\alpha = (\alpha_k)_{1 \leq k \leq N-2}$ is given by

$$\alpha = V^{-1}\Big(\mathbf{A}^{-1}\mathbf{B} + \mathbf{f}(0)\Big)$$

By integration,

$$p_N(t) = M\Big[\sum_{k=1}^{N-2} \alpha_k V_{N-2}^{(k)} \frac{e^{\lambda_k t} - 1}{\lambda_k} - (\mathbf{A}^{-1}\mathbf{B})_{N-2} t\Big].$$

Finally, the tail distribution of the nucleation time is given by

$$S_T(t) = \mathbb{P}\big\{T > t\big\} = \mathbb{P}\big\{P_N(s) = 0, 0 \leq s \leq t\big\} = \mathbb{P}\big\{P_N(t) = 0\big\} = \exp(-p_N(t)),$$

the last equality being the consequence that $P_N(t)$ follows a Poisson distribution of mean parameter $p_N(t)$.

Asymptotic We detail below two asymptotics, which are of interest for their own, as well for the original model described in table 3. The two limits we look at are the unfavourable case $q \gg M$ and the favourable case $M \gg q$. In such cases the mean lag time is given by

$$\begin{aligned} &< T > \sim_{M \gg q} \frac{(2(N-1)!)^{1/(N-1)}}{M^{N/(N-1)}}, \\ &< T > \sim_{M \ll q} \frac{2q^{N-2}}{M^N}, \end{aligned} \tag{25}$$

from which we can deduce

$$\begin{aligned} &\log(< T >) \sim_{M \gg q} -\frac{N}{N-1}\log(M) \sim -\log(M), \\ &\log(< T >) \sim_{M \ll q} -N\log(M). \end{aligned} \tag{26}$$

Similarly, there are two different asymptotic distributions for the lag time, given respectively by a Weibull and an exponential distribution,

$$\begin{aligned} &u_T(t) \sim_{t \to 0} \frac{M^N}{2(N-2)!} t^{N-2} \exp\Big(-\frac{M^N}{2(N-1)!} t^{N-1} \Big), \\ &u_T(t) \sim_{t \to \infty} \frac{M^N}{2q^{N-2}} \exp\Big(-\frac{M^N}{2q^{N-2}} t \Big). \end{aligned} \tag{27}$$

Hence, the mean nucleation time in the monomer constant formulation model has two main different behaviours:

- For unfavourable aggregation ($M \ll q$), the nucleation time distribution is exponential, and the mean nucleation time depends strongly on the number of active monomers M, linearly with a factor $-N$ in log scale.
- For favourable aggregation ($M \gg q$), the nucleation time distribution is a Weibull distribution, and the mean nucleation time depends weakly on the number of active monomers M, linearly with a factor almost equal -1 in log scale.

3.3 Numerical methods

We performed a Gillespie algorithm, a stochastic simulation algorithm [26] from the stochastic models described in table 2 and table 3. In these stochastic models, reaction propensities are given through action-mass law (with discrete number of molecules) and state change vector by the stoichiometry of the reactions. The algorithm simulates the successive stochastic discrete events that occur in the system and is an exact formulation of the stochastic process. Figure 2 was obtained from this algorithm for the model described in table 2. For the study of the nucleation time, stochastic simulation were obtained for the model described in table 3 with $M_A(0) = M$, $P_i(0) = 0$ for $2 \leq i$, and $\gamma = \gamma^* = k_b = 0$. Simulation were stopped when $P_N = 1$ and the nucleation time was recorded accordingly. The initial number of active monomers M was taken to vary over 9 orders of magnitude (from 10^0 to 10^9). The equilibrium constant q was taken to vary over similar ranges of magnitude, to investigate both cases $q \ll M$ and $q \gg M$. The nucleation size N was taken to vary from 3 to 20, according to literature [33, 48].

3.4 Limit theorem

Suppose we have $P(t)$ polymers of size respectively R_i, $i = 1..P(t)$, and that each polymers grow by addition of one by one monomers, at a speed given by a Poisson process of parameter $k_p M(t)$ where $M(t)$

is the quantity of monomer at instant t. Assume similarly that each polymer of size R_i can break into two pieces of size $1 \leq r \leq R_i - 1$ and $R_i - r$, following a Poisson process of rate $R_i k_b$. The splitting size r is chosen uniformly among $[1, R_i - 1]$. Now suppose the total mass of monomers is large, of order n a large scaling parameter. We then let

$$\begin{array}{rcl} m^n(t) & = & M/n, \\ r_i^n(t) & = & R_i/n. \end{array} \tag{28}$$

As $n \to \infty$, the standard scaling of poisson process gives us that $(m^n, (r_i^n)_{1 \leq i \leq P(t)})$ converges to the solution of the hybrid model given by

$$\left\{ \begin{array}{rcl} \dot{m}(t) & = & -k_p m(t) P(t), \\ \dot{r}_i(t) & = & +k_p m(t), \quad 1 \leq i \leq P(t), \\ P(t) & = & P(0) + Y\Big(\displaystyle\int_0^t k_b z(s) ds \Big), \\ z(t) & = & z(0) + m(0) - m(t) = \displaystyle\sum_{i=1}^{P(t)} r_i(t). \end{array} \right. \tag{29}$$

In such model, each polymer grows at speed $k_p m(t)$ according to an ordinary differential equation, and new polymers appears at rate $k_b z(t)$, given by a poisson process, where $z(t)$ is the total mass of polymers. Due to the linearity assumed for the fragmentation rate, and the size-independent polymerization rate, individual mass of polymers are not relevant, and the system can be reduced to the three variable $m(t), z(t), P(t)$, as in a deterministic framework (see [33]).

Assuming further a faster fragmentation process than elongation process,

$$\begin{array}{rcl} k_b^n & = & n k_b, \\ k_p^n(t) & = & k_p/n, \\ p^n(t) & = & P(t)/n, \end{array} \tag{30}$$

the system (29) converges as $n \to \infty$ (again by a standard limit theorem) to

$$\left\{ \begin{array}{rcl} \dot{m}(t) & = & -k_p m(t) p(t), \\ \dot{z}(t) & = & +k_p m(t) p(t), \\ \dot{p}(t) & = & +k_b z(t). \end{array} \right. \tag{31}$$

To obtain such deterministic model, the fragmentation rate has been assumed large so that a large number of polymers are created. Upon renormalization, the number of polymers is now represented by a continuous variable. Finally, a Central Limit theorem gives us the following approximation for the total mass of growing polymers

$$\hat{z}^n = z(t) + \frac{1}{\sqrt{n}} \xi(t), \tag{32}$$

where $\xi(t) = \lim_{n \to \infty} \sqrt{n}(z(t) - z^n(t))$, $\rho(t) = \lim_{n \to \infty} \sqrt{n}(p(t) - p^n(t))$ are solution of the linear stochastic differential equation

$$\left\{ \begin{array}{rcl} \xi(t) & = & -k_p \displaystyle\int_0^t p(s) \xi(s) ds + k_p \int_0^t (z(0) + m(0) - z(s)) \rho(s) ds, \\ \rho(t) & = & k_b \displaystyle\int_0^t \xi(s) ds + W\Big(\int_0^t k_b z(s) ds \Big) \end{array} \right. \tag{33}$$

where W is a standard Wiener process.

Acknowledgments

References

1. Martin JB (1999) Molecular basis of the neurodegenative disorders. The New England Journal of Medicine 340: 1970-1981.

2. Morris AM, Watzki MA, Finke RG (2009) Protein aggregation kinetics, mechanism, and curve-fitting: A review of the literature. Biochemica and Biophysica Acta 1794: 375-397.

3. Caughey B, Kocisko DA, Raymond GJ, Lansbury PT (1995) Aggregates of scrapie associated prion protein induce the cell-free conversion of protease-sensitive prion protein to the protease-resistant state. Chemistry and biology 2: 807-817.

4. Cohen F, KM Pan ZH, Baldwin M, Fletterick R, Prusiner S (1994) Structural clues to prion replication. Science 264: 530-1.

5. Prusiner SB (1998) Prions. PNAS 95: 13363-13383.

6. Liautard JP (1991) Are prions misfolded molecular chaperones? FEBS Letters 294: 155-157.

7. Cohen F, Prusiner S (1998) Pathologic conformations of prion proteins. Annu Rev Biochem 67: 793-819.

8. Kelly JW (2000) Mechanisms of amyloidogenesis. nature structural biology 7.

9. Hess S, Lindquist SL, Scheibel T (2007) Alternative assembly pathways of the amyloidogenic yeast prion determinant sup35. EMBO reports VOL 8.

10. Nguyen HD, Hall CK (2004) Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. PNAS 101: 1618016185.

11. Pellarin R, Caflisch A (2006) Interpreting the aggregation kinetics of amyloid peptides. J Mol Biol 360: 882-92.

12. Padrick SB, Miranker AD (2002) Islet amyloid: Phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis. Biochemistry 41: 4694-4703.

13. Powers ET, Powers DL (2008) Mechanisms of protein fibril formation: Nucleated polymerization with competing off-pathway aggregation. Biophysical Journal 94: 379-391.

14. Serio TR, Cashikar AG, Kowal AS, Sawicki GJ, Moslehi JJ, et al. (2000) Nucleated conformational conversion and the replication of conformational information by a prion determinant. Science 289: 1317-1321.

15. Bishop MF, Ferrone FA (1984) Kinetics of nucleation-controlled polymerization. a perturbation treatment for use with a secondary pathway. Biophys J 46: 631-644.

16. Xue WF, Homans SW, Radford SE (2008) Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. PNAS 105: 8926-8931.

17. Xu S, Bevis B, Arnsdorf MF (2001) The assembly of amyloidogenic yeast sup35 as assessed by scanning (atomic) force microscopy: An analogy to linear colloidal aggregation? Biophysical Journal 81: 446-454.

18. Hur K, Kim JI, Choi SI, Choi EK, Carp RI, et al. (2002) The pathogenic mechanisms of prion diseases. Mechanisms of Ageing and Development 123: 1637-1647.

19. Eghiaian F, Daubenfeld T, Quenet Y, van Audenhaege M, Bouin A, et al. (2007) Diversity in prion protein oligomerization pathways results from domain expansion as revealed by hydrogen/deuterium exchange and disulfide linkage. PNAS 104: 74147419.

20. Alvarez-Martinez MT, Fontes P, Zomosa-Signoret V, Arnaud JD, Hingant E, et al. (2011) Dynamics of polymerization shed light on the mechanisms that lead to multiple amyloid structures of the prion protein. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1814: 1305–1317.

21. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75: 333-66.

22. Moore RA, Taubner LM, Priola SA (2009) Prion protein misfolding and disease. Curr Opin Struc Biol 19: 14–22.

23. Priola SA, Vorberg I (2006) Molecular aspects of disease pathogenesis in the transmissible spongiform encephalopathies. Mo biotechnol 33: 71–88.

24. Anderson DF, Kurtz TG (2011) Design and Analysis of Biomolecular Circuits (chapter 1). Springer.

25. G S Jackson APAFHJKHSCJCJPWARCJC L L P Hosszu (1999) Reversible conversion of monomeric human prion protein between native and fibrilogenic conformations. Science 283: 1935-37.

26. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81.

27. Kang HW, Kurtz TG (2013) Separation of time-scales and model reduction for stochastic reaction networks. Ann Appl Probab 23: 529-583.

28. Fenichel N (1979) Geometric singular perturbation theory for ordinary differential equations. J Differ Equations 31: 53–98.

29. Yvinec R, D'Orsogna MR, Chou T (2012) First passage times in homogeneous nucleation and self-assembly. J Chem Phys 137: 244107.

30. Gillespie CS (2009) Moment-closure approximations for mass-action models. IET Syst Biol 3: 52-58.

31. Penrose O (1989) Metastables states for the becker-döring cluster equations. Commun Math Phys 541: 515–541.

32. Wattis JAD, King JR (1998) Asymptotic solutions of the becker-döring equations. J Phys A: Math Gen 31: 7169-7189.

33. Masel J, Jansen V, Nowak M (1999) Quantifying the kinetic parameters of prion replication. Biophysical chemistry 77: 139–52.

34. Knowles TPJ, Waudby CA, Devlin GL, Cohen SIA, Aguzzi A, et al. (2009) An analytical solution to the kinetics of breakable filament assembly. Science 326: 1533.

35. Szabo A (1998) Fluctuations in the polymerization of sickle hemoglobin a simple analytic model. J Mol Biol 539-542: 539-542.

36. Ferreira AS, da Silva MAA, Cressoni1 JC (2003) Stochastic modeling approach to the incubation time of prionic diseases. Phys Rev Lett 90: 198101-1 198101-4.

37. Hofrichter J (1986) Kinetics of sickle hemoglobin polymerization. iii. nucleation rates determined from stochastic fluctuations in polymerization progress curves. J Mol Biol 189: 553-571.

38. Kulkarni R, Slepoy A, Singh R, Cox D, Pzmndi F (2003) Theoretical modeling of prion disease incubation. Biophys J 85: 707-18.

39. Schweitzer F, Schimansky-Geier L, Ebeling W, Ulbricht H (1988) A stochastic approach to nucleation in finite systems: theory and computer simulations. Physica A 150: 261-279.

40. Bhatt JS, Ford IJ (2003) Kinetics of heterogeneous nucleation for low mean cluster populations. J Chem Phys 118: 3166-3176.

41. Goldstein RF, Stryer L (1986) Cooperative polymerization reactions: Analytical approximations, numerical examples, and experimental strategy. BIOPHYS J 50: 583-599.

42. Powers ET, Powers DL (2006) The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the supercritical concentration. Biophysical Journal 91: 122-132.

43. Baskakov IV, Bocharova OV (2005) In vitro conversion of mammalian prion protein into amyloid fibrils displays unusual features. Biochemistry 44: 2339-2348.

44. Ferrone FA (2006) Nucleation: The connections between equilibrium and kinetic behavior. Methods in Enzymology 412.

45. Prigent S CFLNGPea Ballesta A (2012) An efficient kinetic model for assemblies of amyloid fibrils and its application to polyglutamine aggregation. PLoS ONE 7: e43273.

46. Kingman JFC (1969) Markov population processes. J Appl Probab 6: 1-18.

47. Yueh WC (2005) Eigenvalues of several tridiagonal matrices. Applied Mathematics E-Notes 5: 66–74.

48. Masel J, Genoud N, Aguzzi A (2005) Efficient inhibition of prion replication by prp-fc2 suggests that the prion is a prpsc oligomer. Journal of Molecular Biology 345: 1243-1251.

Figure Legends

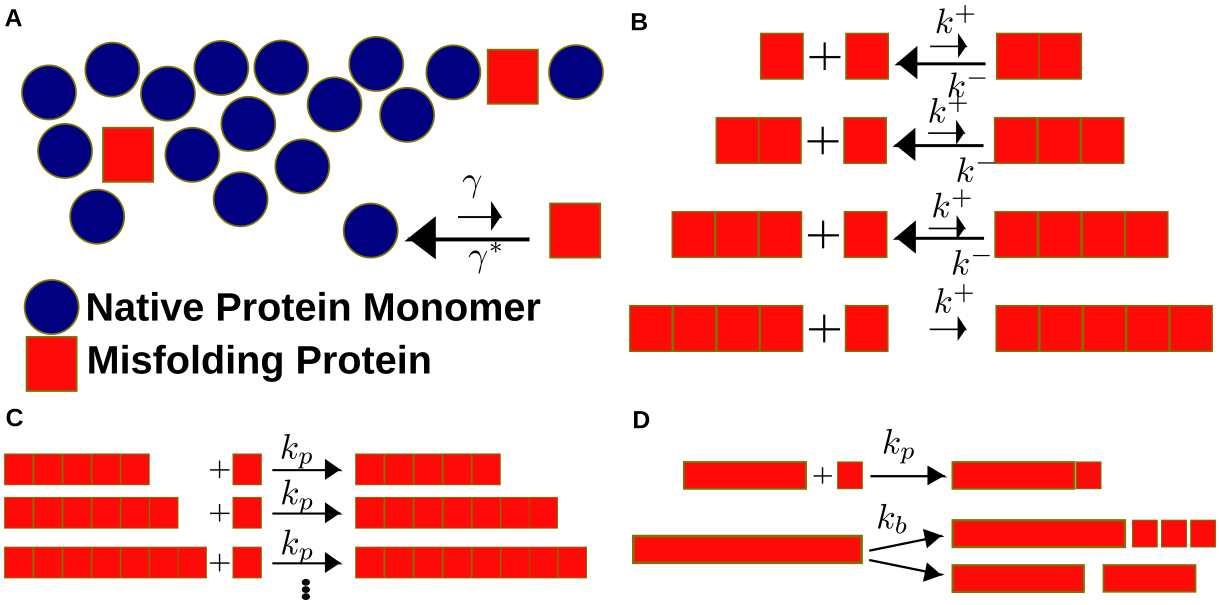

Figure 1. Schematic view of the dynamical model of Prion nucleation-polymerization with misfolding. **A** Spontaneous Misfoling reaction, that consists of reversible first-order kinetic reaction by which monomer changes conformation, at rate respectively γ and γ^*. In the reduced model, we consider that a fast equilibrium between normal and misfolded monomer is reached instantaneously ($\gamma, \gamma^* \to \infty$). **B** Nucleation steps (with $N = 5$). All the steps are composed of reversible lengthening/shortening of a single monomer. The forward aggregation reaction is thus a second-order reaction (with kinetic rate k^+, independent of the size of the aggregate), and the backward disaggregation reaction is a first-order reaction (with kinetic rate k^-, independent of the size of the aggregate). The first nucleation event refers to the first nucleus formed, that is an aggregate of size N. **C** Polymerization steps. For aggregates of size larger than N, the aggregation process is composed of irreversible addition of a single monomer, as a second-order kinetic reaction (with kinetic rate k_p, independent of the size of the aggregate). **D** Polymerization/fragmentation steps. As the polymer becomes larger and larger, the polymerization process still occur and the fragmentation becomes more likely. The fragmentation rate is *proportional* to the size of the polymer, in a first-order kinetic reaction (with kinetic rate k_b). The two parts have equal probability to be of a size between one and the size of the initial polymer minus one. When it gives birth to an oligomer (size less than N) this last one is supposed to break into small monomers immediately due to the instability of the oligomer).

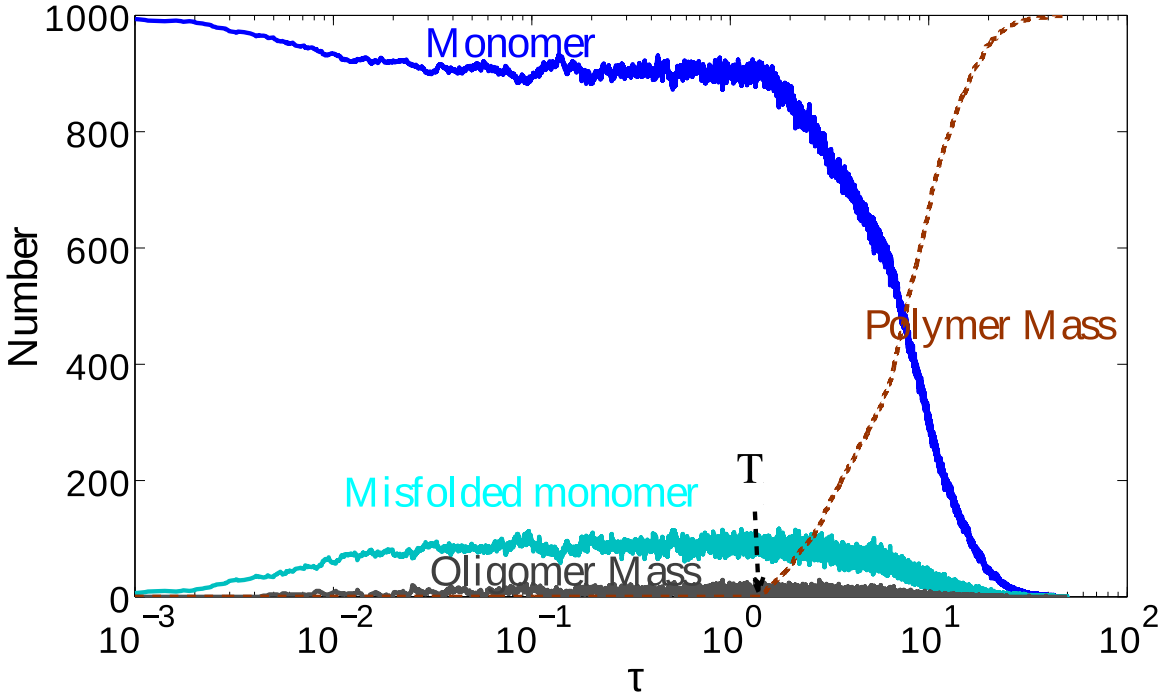

Figure 2. One stochastic simulation of the full stochastic discrete model, and definition of the lag time. One simulation of the stochastic model, with the numbers of normal and misfolded protein, the mass of oligomers (aggregates of size less than N) and the mass of polymers (aggregates of size larger than N). We use here aggregate variables to improve visualization. The lag time is defined as the waiting time for the formation of the first nucleus . We used $M = 1000$, $\gamma^*/\gamma = 10$, $q = 1000$, $N = 7$. The time (in log scale) has been rescaled by $\tau = k^+ t$.

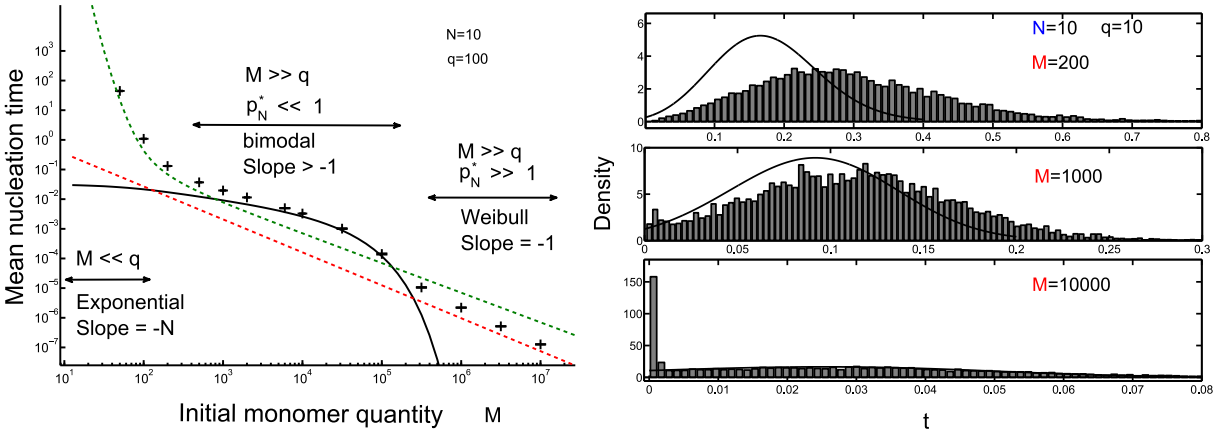

Figure 3. Statistics of the nucleation time in the stochastic nucleation model. **Left Panel:** Mean nucleation time as a function of the initial quantity of monomers. The solid black symbol show the statistical mean obtained through numerical result (for 10000 runs). The red dashed line is given by the Weibull approximation, the green dashed line is given by the exponential approximation, and the solid black line by the linear metastable approximation. Parameter are $q = 100$, and $N = 10$ as indicated on the figure. **Right Panel:** Nucleation distribution time, for $N = 10$, $q = 10$ and from top to down, $M = 200, 1000, 10000$. The histogram show the numerical result (for 10000 runs), while the solid black line is given by the linear metastable approximation.

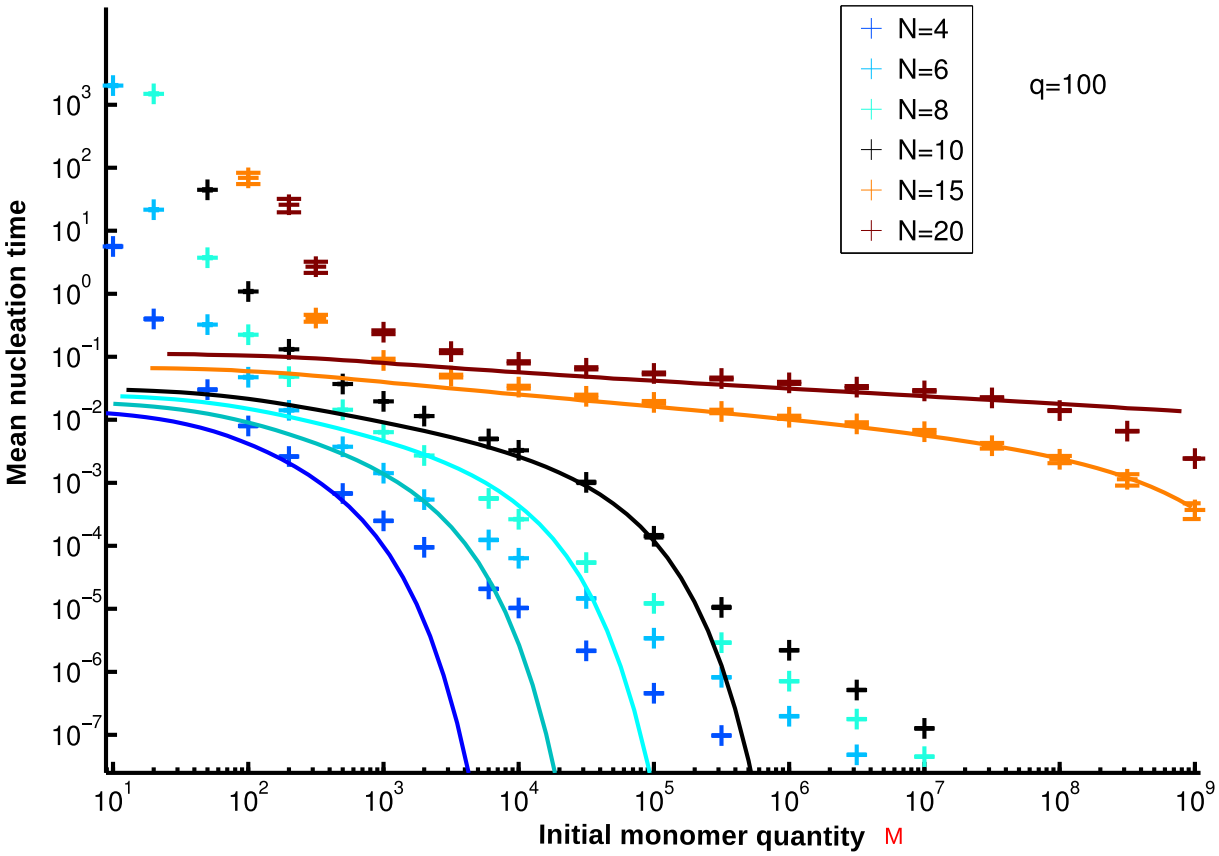

Figure 4. Mean nucleation time in the stochastic nucleation model. Mean nucleation time as a function of the initial quantity of monomer. The solid color symbol show the statistical mean obtained through numerical result (for 10000 runs), while the solid color line are given by the linear metastable approximation. Each color corresponds to a nucleus size of $N \in [4, 20]$, as indicated on the legend. Parameter $q = 100$.

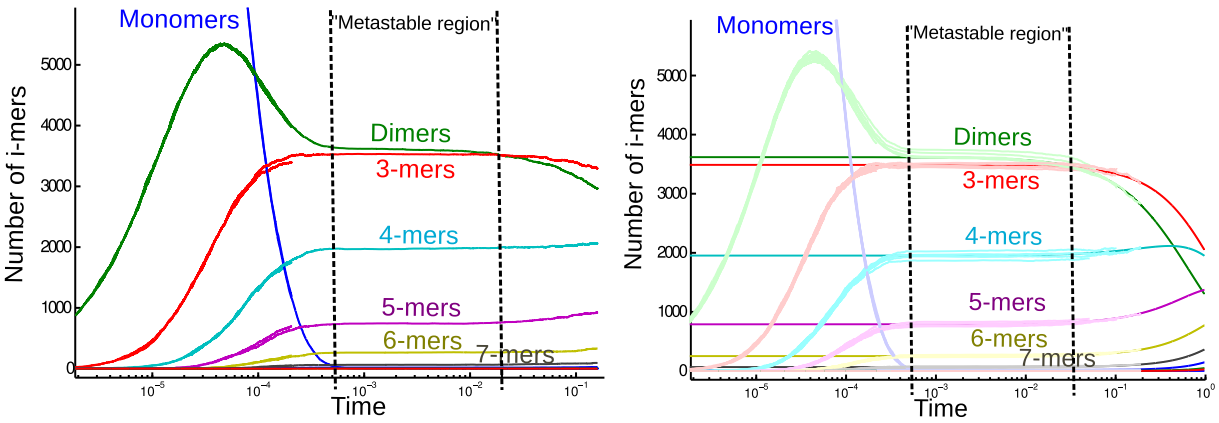

Figure 5. Metastable trajectories in the stochastic nucleation model. $M = 30000$, $q = 1$, $N = 10$. Each solid color line represents the time evolution of the number of i-mer, as indicated on the legend. **Left Panel:** Two different trajectories are shown. In the first, a nucleus is formed at time $\sim 10^{-4}$, and in the second one at time $\sim 10^{-1}$. **Right Panel:** Illustration of the metastable linear approximation. The solid color lines, superimposed on several simulation trajectories, are the result of the linear metastable approximation: initial values are given by the metastable values p_i^* and the numbers of aggregates evolves following the linear model.

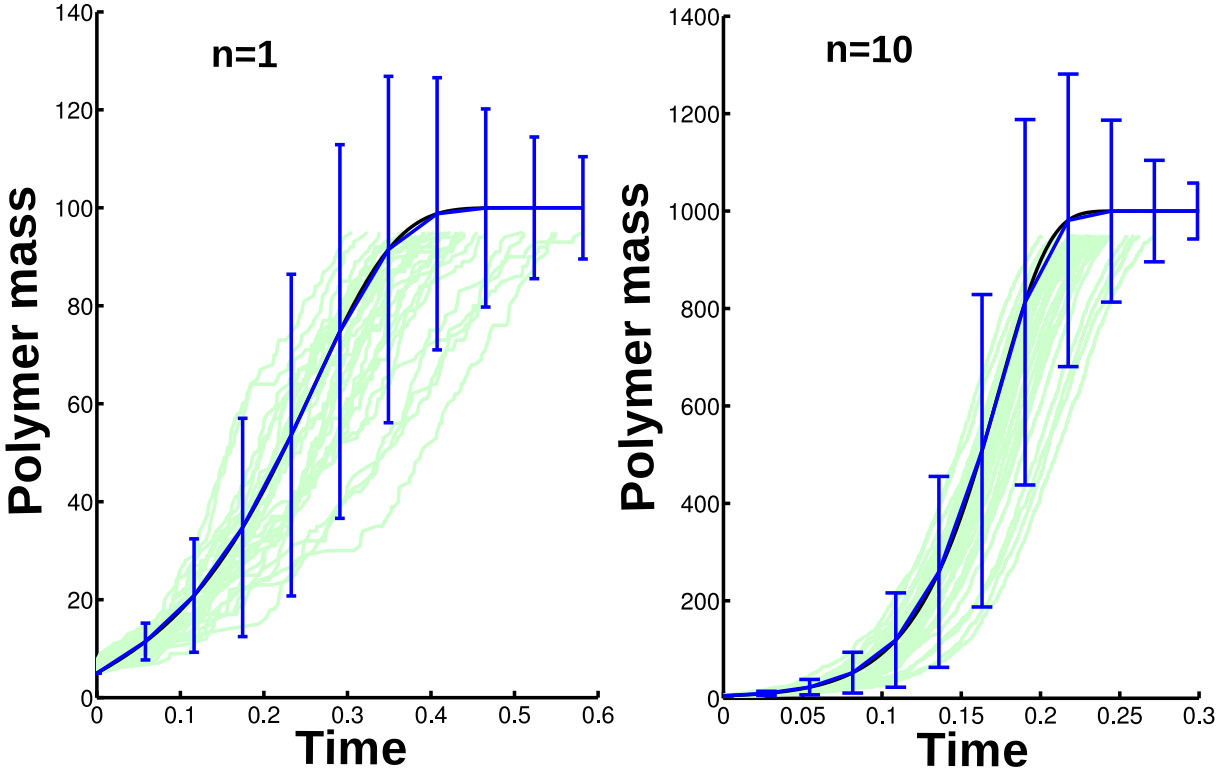

Figure 6. Simulations of the nucleation-polymerization model in the unfavourable limit $q \gg M$. The solid blue line show the Mean and variance evolution of the mass of polymer, given by the hybrid approximation calculated in the results section 1.4.1. Green lines show several simulation trajectories of the nucleation-polymerization model, with $M = 100$ (**left panel**) and $M = 1000$ (**right panel**). Trajectories have been shifted to coincide at their tenth percent value.

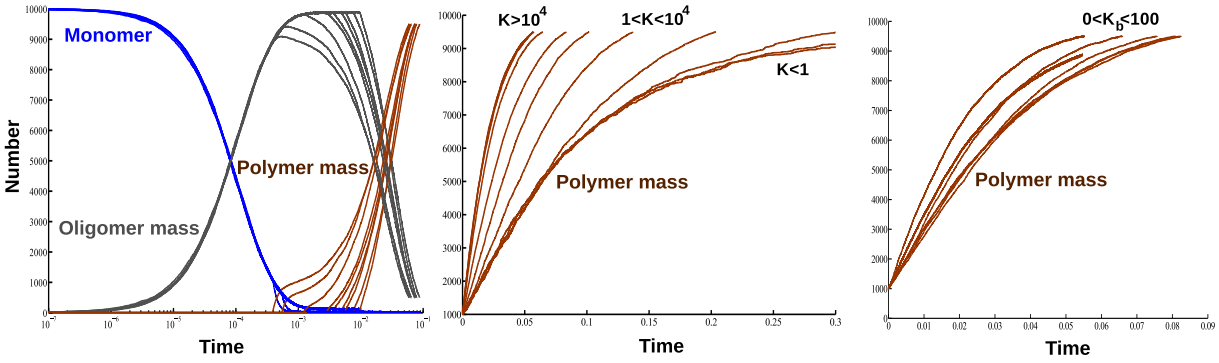

Figure 7. Simulations of the nucleation-polymerization model in the nucleation dominant limit $M \gg q$. Parameters are $M = 1000$, $q = 10$, $N = 10$. **Left Panel:** Several spontaneous polymerization simulations are shown. We observe that oligomers are created in large quantity before nucleation occurs. The nucleation time is highly variable from one simulation to another, but polymerization speed is similar from one simulation to another and governed by oligomer disappearance. textbfMiddle Panel: Several polymers mass time evolution are shown, shifted to coincide at their tenth percent values. Different values of the polymerization rate K_p are used. This parameter have little effect if larger to a given values (10^4 here). **Right Panel:** Several polymers mass time evolution are shown, shifted to coincide at their tenth percent values. Different values of the fragmentation rate K_b are used. This parameter have little effect on polymerization speed.

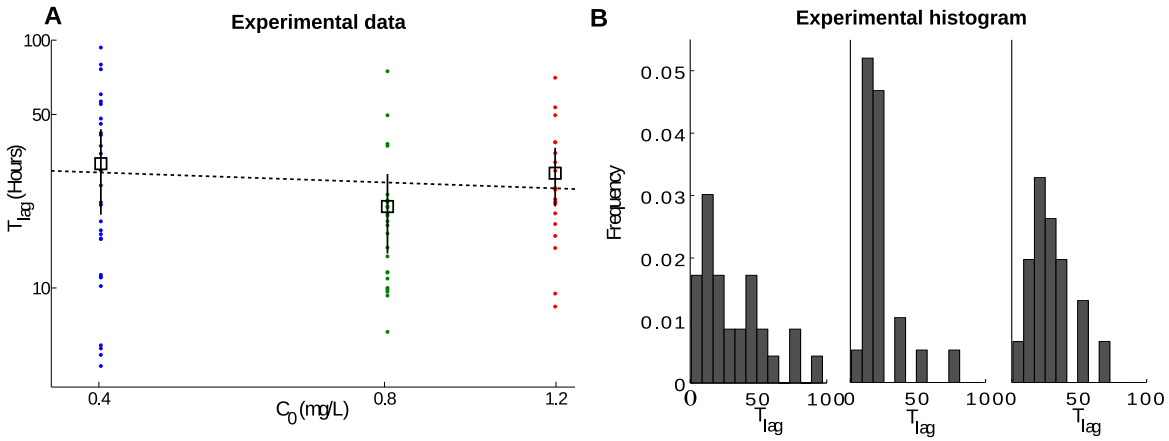

Figure 8. Experimental data of nucleation time statistics from spontaneous prion polymerization experiments, extracted from [20]. In the **left panel A**, each color (respectively blue, green, red) symbol represent the nucleation time in a single spontaneous prion polymerization experiment, for a given initial prion protein PrP^c concentration (respectively 0.4, 0.8 and 1.2 mg/L). Square symbol represent the mean nucleation time for a given initial prion protein concentration over the repeated experiments, and the vertical line represents the normalized variance. The dashed line is obtained by a linear fit of the means as a function of initial concentration (in log scale). The slope is $-0.13\ hours^{-1}.mg^{-1}.L$. The correlation coefficient between the nucleation time and the initial concentration is -0.08, with a p-value of 0.49. In the **right panel B**, the same experimental data are represented as histograms. From left to right, initial prion protein PrP^c concentrations are respectively 0.4, 0.8 and 1.2 mg/L. The histograms are constructed based on the points of the left panel, with respectively 29, 24 and 19 experiments.

Tables

Table 1. Definitions of variables and parameters in the full stochastic nucleation-polymerization model.

Symbol	Definition
M_I	Number of inactive native monomer
M_A	Number of active misfolded monomer
M	Total initial monomer quantity
P_i	Number of polymer of size $i \geq 2$
N	Nucleus size
γ	Folding rate
γ^*	Unfolding rate
$c_0 = \frac{\gamma^*}{\gamma}$	Equilibrium constant between monomers
k^+	Elongation rate in nucleation steps
k^-	Dissociation rate in nucleation steps
$q = \frac{k^-}{k^+}$	Equilibrium constant in nucleation steps
k_p	Elongation rate in polymerization steps
k_b	Fragmentation rate in polymerization steps
q_0	Normalized constant $q(1 + c_0)$
K_b	Normalized constant $\frac{k_b}{k^+}(1 + c_0)$
K_p	Normalized constant $\frac{k_p}{k^+}(1 + c_0)$

We sum up in this table the variable and parameters involved in the full stochastic discrete model of nucleation-polymerization. One variable is used for the number of each species: M_I, M_A and P_i for every $i \geq 2$. For each kinetic reaction described in the model, we use a single kinetic rate constant (mass-action law): γ, γ^*, k^+, k^-, k_p and k_b. The other parameters are the nucleus size N at which the reaction scheme changes, and the initial condition given the total number of monomers M. Equilibrium constant are denoted by c_0 and q. The remaining constant q_0, K_b and K_p are normalized constant as indicated in the table. See also section 1.1 for details.

Table 2. Propensity and state-transition of each reaction in the full stochastic nucleation-polymerization model.

Reaction	State transition	Propensity
Folding reaction	$M_I \to M_I - 1, M_A \to M_A + 1$	γM_I
Unfolding reaction	$M_I \to M_I + 1, M_A \to M_A - 1$	$\gamma^* M_A$
Dimer formation	$M_A \to M_A - 2, P_2 \to P_2 + 1$	$k_+ M_A(M_A - 1)/2$
i-mer formation ($N \geq i \geq 3$)	$M_A \to M_A - 1, P_{i-1} \to P_{i-1} - 1, P_i \to P_i + 1$	$k_+ M_A P_{i-1}$
i-mer destruction ($N - 1 \geq i \geq 2$)	$M_A \to M_A + 1, P_{i-1} \to P_{i-1} + 1, P_i \to P_i - 1$	$k_- P_i$
Polymerization ($i \geq N + 1$)	$M_A \to M_A - 1, P_{i-1} \to P_{i-1} - 1, P_i \to P_i + 1$	$k_p M_A P_{i-1}$
Fragmentation ($i \geq N$, $j < i$)	$P_i \to P_i - 1, P_j \to P_j + 1, P_{i-j} \to P_{i-j} + 1$	$2k_b P_i$

We sum up in this table the propensity and state-transition of each reaction involved in the full stochastic discrete nucleation-polymerization model (see also Figure 1). The state-transition of each reaction is given by the stoichiometry of the biochemical reaction, and the propensity is given by mass-action law, with a given kinetic rate constant. This uniquely defines a time-continuous Markov chain on a discrete state-space. See section 1.1 for details.

Table 3. Propensity and state-transition of each reaction in the reduced stochastic model with fast misfolding process (conditions $(H1)$) and rescaled parameters.

Reaction	State transition	propensity
Dimer formation	$M_F \to M_F - 2, P_2 \to P_2 + 1$	$\frac{1}{2(1+c_0)} M_F(M_F - 1)$
i-mer formation $(N \geq i \geq 3)$	$M_F \to M_F - 1, P_{i-1} \to P_{i-1} - 1, P_i \to P_{i+1} + 1$	$M_F P_{i-1}$
i-mer destruction $(N - 1 \geq i \geq 2)$	$M_F \to M_F + 1, P_{i-1} \to P_{i-1} + 1, P_i \to P_i - 1$	$q_0 P_i$
Polymerization $(i \geq N + 1)$	$M_F \to M_F - 1, P_{i-1} \to P_{i-1} - 1, P_i \to P_i + 1$	$K_p M_F P_{i-1}$
Fragmentation $(i \geq N,\, j < i\ )$	$P_i \to P_i - 1, P_j \to P_j + 1, P_{i-j} \to P_{i-j} + 1$	$2K_b P_i$

We sum up in this table the propensity and state-transition of each reaction involved in the reduced stochastic model with fast misfolding process (conditions $(H1)$) and rescaled parameters. The state-transition of each reaction is given by the stoichiometry of the biochemical reaction, and the propensity is given by mass-action law, with a proper time renormalization. This uniquely defines a time-continuous Markov chain on a discrete state-space. See section 1.2 for details.

Table 4. Analytical approximation of the nucleation time.

Parameter conditions	Distribution	Mean
$q \gg M$	Exponential	$< M_F P_{n-1} > (t \to \infty) \sim \frac{2q^{N-2}}{M^N}$
$M \gg q$, $p_N^* \gg 1$	Weibull	$\frac{(2(N-1)!)^{1/(N-1)}}{M^{N/(N-1)}}$
$M \gg q$, $p_N^* \ll 1$	Bimodal	nearly independent of M

In this table, we sum up the different analytical approximations of the nucleation time, for the full stochastic discrete model of nucleation. There are three different approximations, for the unfavourable case and the favourable case with small or large nucleus size. In the last bimodal case, we can obtain an analytical formula which has not been reported here as no close-form is available (see also Figure 3 and 4). See text in subsection 1.3 for more details.

Table 5. Normalized metastable values $c_i^* = p_i^*/M$ for the nucleation model in the favourable case $M \gg q$.

size	value	size	value
c_2^*	0.1145	c_9^*	9.403910^{-5}
c_3^*	0.1104	c_{10}^*	1.646310^{-5}
c_4^*	0.0618	c_{11}^*	2.591910^{-6}
c_5^*	0.0250	c_{12}^*	3.707610^{-7}
c_6^*	0.0080	c_{13}^*	4.859610^{-8}
c_7^*	0.0021	c_{14}^*	5.878110^{-9}
c_8^*	4.768810^{-4}	c_{15}^*	6.600910^{-10}

In this table, we compute the numerical values of the normalized metastable values $c_i^* = p_i^*/M$ for the nucleation model in the favourable case $M \gg q$. Such values represent the level that each variable reach during the metastable period after the pure-aggregation period (see also Figure 5). See text in subsection 1.3 and 3.1 for more details