**International Journal of Pure and Applied Mathematics**
**Volume 93 No. 6 2014, 845-878**
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
**url:** http://www.ijpam.eu
**doi:** http://dx.doi.org/10.12732/ijpam.v93i6.10


# SIZE DISTRIBUTION OF AMYLOID FIBRILS. MATHEMATICAL MODELS AND EXPERIMENTAL DATA

S. Prigent[1] §, H.W. Haffaf[2], H.T. Banks[3],
M. Hoffmann[4], H. Rezaei[5], M. Doumic[6]

[1,2,6]Inria, Institut National de Recherche en Informatique et Automatique
Rocquencourt, FRANCE

[1,2,6]Pierre et Marie Curie University
Paris-Diderot University
CNRS UMR 7598, Paris, FRANCE

[3]Center for Research in Scientific Computation (CRSC)
North Carolina State University
Raleigh, N.C., USA

[4]CEREMADE (Centre de REcherche en MAthmatique de la DEcision)
CNRS-UMR 7534 and CREST
University Paris-Dauphine
Paris, FRANCE

[5]Institut National de Recherche Agronomique
Jouy-en-Josas, FRANCE

**Abstract:** More than twenty types of proteins can adopt misfolded conformations, which can co-aggregate into amyloid fibrils, and are related to pathologies such as Alzheimers disease. This article surveys mathematical models for aggregation chain reactions, and discuss the ability to use them to understand amyloid distributions. Numerous reactions have been proposed to play a role in their aggregation kinetics, though the relative importance of each reaction

Received: April 22, 2014



§Correspondence author

in vivo is unclear: these include activation steps, with nucleation compared to initiation, disaggregation steps, with depolymerization compared to fragmentation, and additional processes such as filament coalescence or secondary nucleation. We have statistically analysed the shape of the size distribution of prion fibrils, with the specific example of truncated data due to the experimental technique (electron microscopy). A model of polymerization and depolymerization succeeds in explaining this distribution. It is a very plausible scheme though, as evidenced in the review of other mathematical models, other types of reactions could also give rise to the same type of distributions.

**AMS Subject Classification:** 92D25, 34E99, 62G07, 62P10.
**Key Words:** protein aggregation, PrP fiber, Becker-Döring system, statistical test, kernel density estimation

## 1. Introduction

Among the more than 20 proteins associated with amyloid diseases [43] including Alzheimers disease, prion PrP was the first protein considered as a non conventional transmissible agent. The peculiar infectious property of these agents resides in their ability to transmit an infection though devoid of nucleic acid (DNA, RNA) [1]. In contrast to classic pathogens such as viruses, the replication mechanism namely the protein only mechanism does not induce the synthesis of new molecules, but the structure modification of an existing protein, e.g., PrPc for "cellular prion" that is ubiquitously expressed in the human body, into a misfolded structure prone to aggregation (e.g., PrPsc for scrapie).

The infectivity of the modified prion structure is responsible for transmitted encephalopathies such as vCJD or variant Creutzfeldt-Jakob disease, the disease strongly linked to bovine spongiform encephalopathy (BSE) responsible for mad cow disease. The number of human British cases deceased because of vCJD reached a peak in the year 2000 [47]. Nevertheless the risk of two additional types of cases may be envisaged. A second wave of cases could be due to i) a slower disease development in certain individuals (due to a different PrP genetic sequence and/or secondary routes of body infection) and/or ii) an iatrogenic person-to-person spread. Recent experimental and clinical data might corroborate these hypotheses. An experimental study suggests that the protocols of detection previously employed to detect abnormal PrP may have underestimated the human species barrier against BSE prions when tests have been exclusively restricted to brain analysis [8]. An analysis not on brain tissue, but on appendix tissue, indicates a high prevalence (1:2,000) of abnormal

PrP in the present British population [22]. As the authors pointed out, it is not known whether these abnormal- PrP carriers will develop a clinical prion disease within their lifetime and whether they represent a potential source of disease transmission through medical tools or blood and other tissue donation. However, as demonstrated on animals, the risk of iatrogenic transmission of spongiform encephalopathies (TSE), especially through blood transfusion [3], has to be considered.

Prion structure modification is responsible not only for transmitted encephalopathies but also for hereditary diseases such as Creutzfeldt-Jakob disease (CJD), Gerstmann-Strussler-Scheinker syndrome (GSS) and fatal familial insomnia (FFI). No efficient therapy has yet been developed against these diseases. One possible therapy is to target the pathogenic molecules, although a controversy resides in the nature of this: are the final aggregated step, fibrils, a source of infectious transmissible protein (by releasing reactive seeds) or to the contrary, a way for cells to safely isolate abnormal proteins? This controversy has led to two different types of therapeutic approaches, one targeting the monomeric protein units whereas another aims at accelerating fibrillogenesis [29]. The hypothesis of a protective role of amyloid fibrils against disease progression has emerged in parallel with the discovery of functional amyloid or amyloid-like fibrils. Both detrimental and advantageous properties of amyloid or amyloid-like fibrils were described in the animal, fungi, plant and bacteria kingdoms, e.g., a structural template for melanin polymers in mammals, advantageous phenotypes for yeast and biofilms for bacteria [21, 32, 41]. Their mechanical properties also make amyloid-like fibrils interesting candidates for future nanomaterial.

The formation of these amyloid or amyloid-like fibrils follows kinetics which is governed by molecular mechanisms of protein aggregation and disaggregation. Though this kinetics is usually summarized as a nucleation-elongation process, it is not clear whether nucleation and/or monomer activation occurs. De novo fibril formation is usually considered to be due to a hypothetical primary nucleation where basic units, monomers, would form a nucleus of a critical size which will then develop into filaments, precursors of mature fibrils. In fibril forming systems where no oligomeric species is experimentally detected as a critical on-pathway intermediate for filament growth, monomer addition is usually assumed to be the dominant elongation process. Such elongation by monomer addition was for instance demonstrated for the growth of amyloid fibril of yeast prion Sup35 [13]. In the opposite direction to the elongation by monomer addition (or by addition of oligomeric species), depolymerization, i.e., dissociation occurs by the loss of monomers (or oligomeric species).

Though the number of polymers may be increased by fragmentation of protein aggregates, the relative contribution of fragmentation to the formation of additional polymers remains to be quantified. The occurence of fragmentation may be enhanced in vivo by molecular partners (chaperones or heat shock proteins) or mechanical stress (membrane interaction) as indicated by in vitro experiments on mammal prion (PrP) fibrils [45]. The stability of the polymers against dissociation which includes depolymerization, by loss of monomers, and fragmentation may depend on the maturity level of the fibrils, mature fibrils being more stable than early aggregates against high pressure or high temperature treatments [30]. However, the level of maturity may not affect stability against chemical treatment, as early and late aggregates can be disrupted by chemical treatments with equal efficiency. As discussed by the authors [30], stability highly depends on the nature of the solution (e.g., pH) used to form fibrils or to test fibril stability as environment controls protein structure. The same amyloid protein, for instance A$\beta$ $1-40$ involved in Alzheimers disease, can lead to polymorphism, i.e., several monomer conformations and filament packing, depending on the aggregation conditions [38]. Differences in fibril stability are thus related to the versatility in the structure of the assemblies [46].

In addition to primary nucleation, elongation (polymerization), dissociation (depolymerization) and fragmentation, amyloid kinetics is further influenced by additional processes that have been recently reviewed by Gillam and MacPhee [23]: monomer activation, formation of prenuclear species, secondary nucleation of monomers on fibril surface and branching, association of two polymers (described as "coagulation" in mathematics or "coalescence" in biology) including lateral association and bundling, conformational rearrangement (including micelles, i.e., loop closure for some amyloid proteins and an hypothetical lock-and-dock mechanism) and off-pathways. As some of these processes depend on polymer size, maturity level and/or concentrations, modelling the whole scheme of the reactions that govern kinetics of fibril growth is quite complex.

Prior to the use of fibrils for clarifying amyloid mechanisms or for applied goals, it is necessary to determine well the biophysical parameters of the fibrils. Among these parameters, the measurement of fibril size is a basic requirement that permits, for example. one to determine to what extent fibrils could be a source of infectious protein, by releasing reactive seeds. To focus on size distribution, we have measured the size distribution of prion fibrils from experimental data and compared it to mathematical modelling of size distribution. Our mathematical approach analyses two possible cases. The first case, by using the total polymerized mass as a basis, corresponds to a state where both formation

of new fibrils and elongation of fibrils take place. The second case considers that no additional fibril is formed and thus it can be applied to a biological state where elongation predominates and no new fibril formation occurs. In both cases, dissociation, i.e., depolymerization is taken into account. We show that an exponential law describes well the size distribution of prion fibrils from our experimental data. We further confirm this exponential trend by both mathematical models obtained at equilibrium. This article also provides guidelines to experimentalists for statistical analysis of their own data. This may be of particular interest for experimental techniques which have censored data, i.e., have access only to a part of the values, such as our results from microscopy lacking access to the smallest polymer sizes. Our presentation may also assist biophysicists or biologists in the choice of the minimal number of objects (polymers) that have to be measured to obtain a statistically representative size distribution for polymers and to determine the underlying mathematical laws (exponential for instance). Overall this paper aims at better characterizing the size distribution of fibrils, the size distribution being information that is useful for the analysis of biological mechanisms such as the amyloid biochemistry and for the development of potential nanomaterial.

After describing the experimental procedures performed to measure fibril sizes at a state that is supposed to be near equilibrium in Section 2, the shape of the size distribution is statistically analysed in Section 3. Then in Section 4 a mathematical model of polymerization-depolymerization is developed which is in agreement with our experimental data. The possibility of additional or different reactions (nucleation vs. initiation, fragmentation vs. depolymerization, secondary nucleation) is then discussed in Section 5 along with a literature review of existing mathematical models.

## 2. Experimental Procedures to Measure Fibril Sizes

### Formation of Fibrils. Experimental Methods

Purified fibrillar aggregates from recombinant human PrP monomers obtained from the method of Rezaei [40] were prepared as described in El Moustaine et al., [19]. Briefly, to obtain a large amount of fibrillar aggregates within one week, PrP monomers (22 $\mu$M) were incubated in a classic protein-denaturing solvent (guanidinium chloride) at 310 K (37°C) with a slow rotating agitation (22 rpm on a 1 cm-radius axis, i.e., 0.005 g). The polymerized mass formation

was estimated on aliquots by thioflavin T fluorescence (from 460 up to 520 nm after excitation at 445 nm) until reaching a plateau for fluorescence. The fibrillar aggregates were then concentrated and separated from monomers and small oligomers by three steps of centrifugation (at 17,900 g during 45 min for each step). At each step the fibrillar aggregates were suspended and washed in sodium acetate buffer (10 mM, pH 5.0 using acetic acid for pH adjustment). The concentration of fibrillar aggregates was calculated from absorbance at 280 nm after baseline correction for turbidity and the samples were stored at 277 K (4°C).

### Estimation of the Experimental Number of Monomers per Fibril

We collected two samples, one of size $n = 531$ and one of size $n = 95$. The frequency of the size expressed as monomers per fibril is presented on Figure 1.

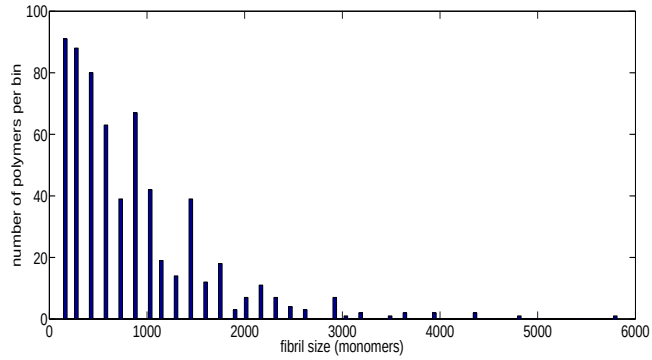


Figure 1: Experimental distribution of the sample $x_i$, $1 \leq i \leq n$, representing the measured sizes of polymers (the number of polymers below a certain size (145 monomers) is unknown). The total size of the sample is $n = 626$, resulting from the combination of two samples of respective sizes $n = 531$ and $n = 95$. For instance, $x_i = 145$ is satisfied for 91 values of $i$ on this sample, which gives a bar of height 91 for $x = 145$.

The size of the each fibril was measured after dilution (to 0.1 $\mu$M) of aliquots from the stock solution of fibrils, by transmission electron microscopy (after negative staining by uranyl acetate on Formvar carbon-coated grids as described in El Moustaine and co-workers [19]). From these images, the length and width

(in nm) were determined. The number of monomers per fibril was roughly estimated from the measured fibril volume divided by a theoretical numerical value that represents the volume of one molecule of native PrP. The PrP included in fibrils, enriched in $\beta$ sheet structure, is certainly packed into a different volume than the native PrP that mainly presents an alpha helix structure. However, the fine structure of the PrP included in fibrils and, therefore, the volume of the fibrillar PrP are not known. For this reason we used the theoretical value obtained for the well-known fine structure of the native PrP (28 $nm^3$ per molecule of PrP) [2], to roughly characterize the overall fibril size distribution expressed as an estimation of monomers per fibril.

The minimum size that can be detected by microscopy corresponds to roughly 145 monomers. Two types of noise affect the values of size determination:

i) a lack of technical precision for the measurement in nanometers,

ii) the uncertainty on the conversion from the measured volume in $nm^3$ to a size in number of monomers.

Adding these two types of noise leads to an uncertainty from $-53$ to $+73$ monomers per polymer. These values were calculated as follows.

We suppose the volume of one molecule of fibrillar PrP is $28 \pm 2\ nm^3$. Furthermore the minimum detectable size on our MET images was measured as $34.8 \pm 1.0\ nm$. From the volume formula of the number of monomers per polymer which is considered as a cylinder, we have

$$\text{number of monomers} = \frac{\text{measured polymer length in nm} \times \pi \times r^2}{\text{monomer volume}}.$$

With the radius $r = 6.1 \pm 1.0nm$, the smallest detectable polymer size is contained between

$$\frac{(34.8 - 1.0)\pi(6.1 - 1.0)^2}{28 + 2} = 92$$

and

$$\frac{(34.8 + 1.0)\pi(6.1 + 1.0)^2}{28 - 2} = 218.$$

As the smallest detectable polymer is considered to be roughly 145 monomers (from $34.8 \times \pi \times 6.1^2/28$ ), this leads to a precision around $-53$ and $+73$ monomers per polymer which is an absolute error range imposed by our measurement technology.

## 3. Statistical Analysis of the Data

As exemplified below by the Becker-Döring model, many mathematical models which may describe the formation of fibrils from monomers tend to a steady distribution for the sizes of the fibrils. This is true not only for Becker-Döring equation [5], but also for growth fragmentation equations [15, 4], for stochastic models [16], etc. This is one of the reasons that makes size distribution analysis so important: if taken at an instant where steady state or steady growth is observed, it is possible to interpret it as solution of a time-independent equation and thus to estimate kinetic coefficients from its shape. One key example concerns the growth-fragmentation equation: from the size distribution it is possible to estimate the division rate of the individuals [18, 17, 16].

Based on these previous studies, we can make the assumption that our size sample $x_1, \cdots, x_n$, as depicted in Figure 1, is the realization of $n$ independent identically distributed (i.i.d.) variables, whose size distribution is the steady state (or equivalently the invariant probability measure) of the system of equations/stochastic processes which adequately describe the reaction scheme.

In the case under study, our knowledge of the reactions that truly occur is too limited to choose with certainty among all the available mathematical models. A heuristic approach thus consists first in comparing the experimental empirical distribution with a classical parametric distribution, such as exponential, Weibull, normal etc. If such a parametric distribution proved to be in good agreement with the empirical one, this could permit one to select a model (perhaps not a unique one) able to produce such a distribution. This is the fundamental premise underlying our discussions.

Under a certain threshold $x_{min} = 145$, we are under the level of detection of the experimental technique (MET) so that it is not possible to measure the data. In our sample (see Figure 1), this threshold is equal to the smallest size class where there is a high amount of data. Nevertheless, as the monomers and the soluble oligomers, i.e., soluble small polymers were experimentally separated from the fibrils (by the centrifugation step), the number of the smallest polymers that electron microscopy is not able to detect should tend towards 0 when the size of the polymers tends to 0.

For reasons explained below, in all the following we define the sample of i.i.d. variables $y_i$ as

$$y_i = x_i - x_{min}, \qquad 1 \leq i \leq n, \tag{1}$$

with $x_{min} = 145$ the threshold under which we are not able to measure any fibril. We study this new sample $\{y_i\}_{1 \leq i \leq n}$, which is of course equivalent to a study

of $\{x_i\}_{1\leq i\leq n}$, since the probability density function of $\{y_i\}_{1\leq i\leq n}$ at a certain point $y$ is the probability density function of $\{x_i\}_{1\leq i\leq n}$ taken in $x = y + x_{min}$.

### 3.1. Kernel Density Estimation

Let us assume that we have a sample $y_1, \cdots, y_n$ of $n$ experimental measurements, each $y_i$ being the realization of independent identically distributed random variables, whose underlying size distribution (or *density*) is $N(y)$. This modelling assumption is partly justified in our case by the mathematical analysis on the Becker-Döring system and the large amount of polymers, leading to an equilibrium. The sample can then be considered *as if* the polymers would be independent.

To estimate the underlying density $f(y)$, kernel estimation methods [42] we propose to define an estimate $\hat{f}$ of $f$ by setting

$$\hat{f}(y) := \frac{1}{n}\sum_{i=1}^{n} K_h * \delta_{y_i}(y) = \frac{1}{n}\sum_{i=1}^{n} K_h(y - y_i), \tag{2}$$

where $\delta_{y_i}$ denotes the Dirac mass in $y_i$ and $K_h(y) := \frac{1}{h}K(\frac{y}{h})$ is a mollifier sequence, $K$ being a function satisfying

$$K \in C_0^\infty(\mathbb{R}), \qquad \int K(y)dy = 1.$$

For instance, $K$ may be chosen as the Gaussian kernel. To choose the small parameter $h$, called the "bandwidth" or sometimes the "regularization parameter", there is a trade-off between smoothness of the estimate $\hat{f}$ and accuracy of the estimation: the smaller $h$, the smaller the distance between the empirical distribution $\frac{1}{n}\sum_{i=1}^{n}\delta_{y_i}$ and $\hat{f}$, but the less smooth the estimate $\hat{f}$. Data-driven methods exist, such as the one very recently introduced by Goldenshluger and Lepski [25, 24, 17], which permit one to choose the optimal parameter $h$.

Here we used the ksdensity procedure of the Matlab statistical toolbox. We obtained a first estimation of the density in Figure 2 (curve in a full black line). As expected visually from the experimental data (the histograms of Figure 1), we obtain a decreasing distribution on the range of measured values. This first result led us to test the sample against an exponential law. We could also have tested other types of distributions such as log-normal, Weibull, etc., but all these distributions are one-peaked whereas our distribution is purely decreasing. When fitted with our data, a one-peak distribution would necessarily have its peak situated for some $y < 0$ (or equivalently $x < x_{min}$, in the $x-$variable of

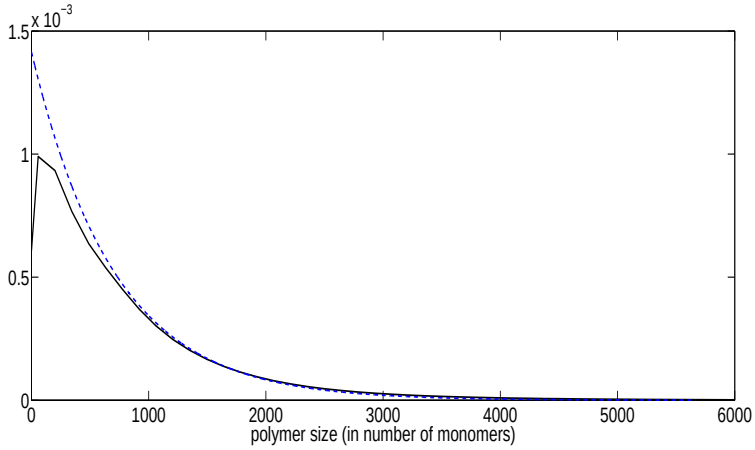


Figure 2: Density estimation of the size distribution of polymers larger than $x_{min}$, represented by $y_i$ (in comparison to Figure 1, all sizes have been translated to the left by a distance of 145 monomers). Dashed line: exponential distribution $p_{\lambda^{MLE}}$, solid line: kernel density estimation.

the initial sample $x_i$, with $x = y + xmin$), what would not provide us with further useful information.

Theory [42] reveals that we expect an approximation of the density on the order of $n^{-\frac{s}{2s+1}}$ where $n$ is the size of the sample and $s = min(p, N)$, $p$ the regularity of the underlying function $f(y)$ ($p-$times differentiable) and $N$ the order of the regularization kernel ($\int y^k K(y)dy = 0$ for $k = 1, \cdots, N$). Even if this is only a qualitative remark, we may take $s = 1$ (for instance for the Gaussian kernel) so that we expect an error in the order of $n^{-\frac{1}{3}}$. This could give clues to determine the size of sample that we need: it is unnecessary to collect a sample that would give a value of $n^{-\frac{1}{3}}$ smaller than the experimental noise. Our dataset (the histograms of Figure 1) shows an average size of 706, and noise is at least on the order of an additive noise of 126 monomers (linked to the precision of the measurement, see Section 2 for the description of the data). This means a noise larger than 18% of the average size value. Evaluating the multiplicative noise linked to the error when converting measured volume to the number of monomers is hazardeous, but we know it is present and should be added to this value. Our sample size is $n = 626$, so that we have $n^{-\frac{1}{3}} \approx 12\%$ which is already smaller than the minimal noise. This implies that it is unnecessary to collect larger samples, as this would not improve our results. This is also

exemplified below: the statistical results on the sample of size 626 are not improved compared to the results obtained for the sample of size 531.

### 3.2. Statistical Tests for Censored Data

Before applying a parametric density distribution $f(y)$ to our data $y_i$, we need to take into account the fact that our measurements will not be distributed along $f(y)$ but rather along $g(y)$ defined by

$$g(y) = \frac{f(y)\mathbb{1}_{y\geq 0}}{\int_0^\infty f(y)dy},$$

with $\mathbb{1}$ representing the Heaviside function: the probability for a measured data to be negative is equal to zero (or, for the original kernel, the probability to be smaller than $x_{min}$).

For the exponential distribution that we want to test here, since it is supported in $\mathbb{R}_+$ we have $g(y) = f(y)$, but this would not be true in general.

#### Maximum Likelihood Estimator

Given the family of exponential laws $p_\lambda := \lambda e^{-\lambda x}$, the maximum likelihood estimation method gives the following closed formula to estimate $\lambda$ from the sample $y_1, \ldots, y_n$ [42]

$$\lambda^{MLE} = \frac{n}{\sum_{i=1}^n y_i}.$$

**Results:** For our data consisting in the concatenation of two samples of the same experiment, one of size $n = 95$ and one of size $n = 531$, we obtain numerically the value $\lambda = 1.415.10^{-3}$ Comparing the value of $\lambda$ for these two samples taken separately, we have an insignificant error smaller than 3% between them, which readily supports combining the data.

### 3.3. $\chi^2$ Test

We would now like to quantify the confidence we may have in the hypothesis $H_0$: "this sample is distributed along $p_{\lambda^{MLE}}$,". This may be done by the $\chi^2$ test, that we describe here briefly for the sake of clarity and for non-statistician readers.

Let us calculate the statistics

$$\chi_n := \sum_{j=1}^{K} \frac{\left(N_j - nP_{\lambda^{MLE}}(A_j)\right)^2}{N_j},$$

where we have divided $\mathbb{R}_+$ in $K$ intervals $A_j$, $1 \leq j \leq K$, and $P_{\lambda^{MLE}}(A_j) = \int_{A_j} p_{\lambda^{MLE}}(s)ds$ represents the probability of the interval $A_j$, and $N_j = \sum_{j=1}^{n} \mathbf{1}_{y_i \in A_j}$ counts empirically the number of data in the class $A_j$.

Theory [48] reveals that for $n \to \infty$, if the hypothesis is true, then $\chi_n$ converges in law to a $\chi^2(K-2)$ distribution. Thus if the distance in law between $\chi_n$ and $\chi^2(K-2)$, for a well-adapted $K$, is sufficiently small, we may accept the hypothesis; otherwise we should reject it.

The hypothesis $H_0$ may be rejected with a confidence $1-\alpha$ if the distance between $\chi_n$ and $\chi^2(K-2)$ is larger than $1-\alpha$. Mathematically speaking, we define the level $q_{1-\alpha}^{K-2}$ as defining the zone where a proportion $\alpha$ of random variables distributed along $\chi^2(K-2)$ would lie: let $Z_{K-2}$ be a random variable of density $\chi^2(K-2)$, $q_{1-\alpha}^{K-2}$ is defined by the relation

$$\mathbb{P}(Z_{K-2} \geq q_{1-\alpha}^{K-2}) = \int_{q_{1-\alpha}^{K-2}}^{\infty} \chi^2(K-2)(s)ds = \alpha,$$

and we can thus define

$$W_{n,\alpha} = \mathbb{1}_{\{\chi_n \geq q_{1-\alpha}^{K-2}\}}.$$

If $W_{n,\alpha} = 0$, we can accept $H_0$; otherwise we reject it.

The largest $\alpha$ such that the hypothesis $H_0$ is accepted is called the p-value, thus defined by

$$p_K = Max\left\{\alpha > 0, W_{n,\alpha} = \mathbb{1}_{\{\chi_n \geq q_{1-\alpha}^{K-2}\}} = 0\right\}.$$

If $p_K$ is sufficiently large, then we are confident in not rejecting the hypothesis $H_0$.

**Results:** Our measurements are discrete and noisy, as shown by Figure 1:

- The precision of the measurement is between $-53$ and $+73$ monomers, so that a minimal size for a bin should be larger.

– Since the distribution is rapidly decreasing and the tested density is exponential, the size of the bins $A_j$ should not be constant but rather determined by a constant value for $nP_{\lambda^{MLE}}(A_j)$.

With these two conditions, since the first bin (sizes around 145 for $x_i$, 0 for $y_i$) contains 91 elements, we propose to define the bins by $nP_{\lambda^{MLE}}(A_j) = 91$. This gives a p-value of 17.6%, what is rather satisfactory, and we may assume that the exponential law is satisfied by our sample - up to the measurement noise described above and the sampling uncertainty. However, due to the measurement errors, the test is sensitive to the number of bins, their position, their average number of data. We may obtain still higher p-values (e.g., for 11 equally-sized bins, we have a p-value of 60%), but also much smaller ones.

In Figure 2 we compare the distribution obtained by a kernel density estimation (through the ksdensity algorithm given by Matlab) and the exponential density $p_{\lambda^{MLE}}$ for the sample $y_i$. Their closeness shows also how much our hypothesis of an exponential distribution may be considered as acceptable. At lowest sizes, the exponential density fits less well than the kernel density estimate, but this is due to boundary effects of the kernel density method, and could also be explained by the truncation of the data.

## 4. Mathematical Model: Polymerization and Depolymerization

To obtain a mathematical transcription of these biological kinetics of the PrP protein formation, two different components can be distinguished: the polymeric fibrils of the PrP protein and the monomeric units of this same protein. The polymeric fibrils (polymers) are considered to be a linear addition of a discrete number of monomeric units (monomers), which represents the fibril size. The kinetic reactions we consider are, hence, two : polymerization and depolymerization. Polymers of size $i$, for $i \geq 1$, can either polymerize (gain one additional monomer) into polymers of size $i+1$, with a coefficient $\mathrm{k}_{\mathrm{on}_i}$, or depolymerize (lose one monomer) into polymers of size $i-1$, with a coefficient $\mathrm{k}_{\mathrm{dep}_i}$. Denoting by $\mathrm{c}_i$ the concentration of polymeric fibrils of size $i$ ($i \geq 1$), we can summarize these reactions in the following scheme :

$$\begin{array}{ccc} \mathrm{c}_i + \mathrm{c}_1 & \xrightarrow{\mathrm{k}_{\mathrm{on}_i}} & \mathrm{c}_{i+1} \\ \mathrm{c}_i & \xrightarrow[\mathrm{k}_{\mathrm{dep}_i}]{} & \mathrm{c}_{i-1} + \mathrm{c}_1 \end{array}$$

Thus, polymers of size $i \geq 1$ can also be a result of polymerization of polymers of size $i-1$ or a result of depolymerization of polymers of size $i+1$. Using the

law of mass action [28] several times, once for each size, we obtain an infinite set of ordinary differential equations, one equation for each size, describing the variation of the concentration of polymers $\mathrm{c}_i$. We arrive at the following system :

$$\begin{cases} \dfrac{d\mathrm{c}_i}{dt} = J_{i-1}(c) - J_i(c) & i \geq 2, \\ \dfrac{d\mathrm{c}_1}{dt} = -\displaystyle\sum_{i=2}^{\infty} J_i(c) - 2J_1(c), \end{cases} \tag{3}$$

where the rate of growth $J_i$ is defined by

$$J_i(c) = \mathrm{k}_{\mathrm{on}_i}\mathrm{c}_1\mathrm{c}_i - \mathrm{k}_{\mathrm{dep}_{i+1}}\mathrm{c}_{i+1} \quad \text{for } i \geq 1. \tag{4}$$

One can observe that one special equation is needed for the monomer variation, because of their interaction with all the polymers.

This infinite system is known as the Becker-Döring system. Introduced in 1935 by Becker and Döring [7], this kinetic model, despite its simple aspects, has been useful to describe various phenomena of phase transition in physics and chemistry (colloid chemistry, nucleation theory, RNA chain formation in the prebiotic world, etc). In their original paper Becker and Döring [7] set the number of monomers, $\mathrm{c}_1$, to be constant in time, which, clearly, does not hold in our experiments. Here the total mass is being conserved just as in the model proposed by Burton [10] and, later, by Penrose and Lebowitz [37].

The existence of positive solutions of (3) has been proved by Ball, Carr and Penrose in [6] under general assumptions (a polymerization coefficient that does not grow faster than the polymer size, i.e., $\mathrm{k}_{\mathrm{on}_i} = O(i), i \to \infty$), while uniqueness is obtained for instance under an additional condition on the second order moment ($\sum_{i=1}^{\infty} i^2\mathrm{c}_i < \infty$). When talking about a solution, we mean a sequence of terms in the functional space

$$\mathrm{E}^+ = \{\mathrm{c} \in \mathrm{E} : \mathrm{c}_i \geq 0 \quad \forall i\},$$

$$\text{where} \qquad \mathrm{E} = \{\mathrm{c} = \{\mathrm{c}_i\}_{i\in\mathbb{N}} : \|\mathrm{c}\| < \infty\}, \quad \|\mathrm{c}\| = \sum_{i=1}^{\infty} i\,|\mathrm{c}_i|.$$

Total mass conservation is then a direct consequence [7, 36]:

$$\rho(t) := \sum_{i=1}^{\infty} i\mathrm{c}_i(t) = \sum_{i=1}^{\infty} i\mathrm{c}_i(0), \qquad t \in [0, T).$$

In the experiments detailed in Section 2, the PrP fibrils are assumed not to polymerize above an upper size $i_{\max}$, so that we consider their polymerization

coefficient to be null so that $\mathrm{k}_{\mathrm{on}_{i\geq i_{\max}}} = 0$. Hence, the existence, uniqueness and mass conservation hold for the solution of our model.

There is a critical mass $\rho_s$ (which can be finite or not) such that for each initial condition $\rho \leq \rho_s$ there exists a unique equilibrium state. If $\rho \geq \rho_s$, then each component $\mathrm{c}_i$ converges to the same equilibrium state component which corresponds to the critical mass $\rho_s$ and the mass excess is lost in the aggregates of infinite mass [6]. The solution $\{\mathrm{c}_i\}_{i\geq 1}$ converges to the equilibrium state strongly in $\mathrm{E}^+$, where an equilibrium state of (3) is given by the sequence $\tilde{\mathrm{c}} = \{\tilde{\mathrm{c}}_i\} \in \mathrm{E}^+$ satisfying

$$J_{i-1}(\tilde{\mathrm{c}}) - J_i(\tilde{\mathrm{c}}) = 0 \qquad i \geq 2, \tag{5}$$

$$\sum_{i=2}^{\infty} J_i(\tilde{\mathrm{c}}) + 2J_1(\tilde{\mathrm{c}}) = 0, \tag{6}$$

with $J_i$ defined by Equation (4). The equation (5) implies that all the growth rates $J_i$ are equal for an equilibrium solution, and the equation (6) implies that $J_1 = 0$ and, hence, $J_i = 0$, $\forall i = 1, 2, ....$ We thus have

$$\mathrm{k}_{\mathrm{on}_i}\tilde{\mathrm{c}}_1\tilde{\mathrm{c}}_i - \mathrm{k}_{\mathrm{dep}_{i+1}}\tilde{\mathrm{c}}_{i+1} = 0, \qquad \forall i = 1, 2, ...$$

By induction, one finds that, for $\mathrm{k}_{\mathrm{dep}_{j+1}} > 0, \forall j$ :

$$\begin{cases} \tilde{\mathrm{c}}_i = Q_i\tilde{\mathrm{c}}_1^i, & i \geq 2 \\ \sum\limits_{i=1}^{\infty} iQ_i\tilde{\mathrm{c}}_1^i = \rho, \end{cases}$$

where $Q_i := \prod\limits_{j=1}^{i-1}\left(\dfrac{\mathrm{k}_{\mathrm{on}_j}}{\mathrm{k}_{\mathrm{dep}_{j+1}}}\right)$ for $i \geq 2$, with $Q_1 := 1$.

For the sake of completeness, we summarize in the following proposition the analytical formulae for $\tilde{\mathrm{c}}_i$ for cases where the coefficients are constant.

**Proposition 4.1** (Steady state profile)**.** *1. Let $k_{on} > 0$ and $k_{dep} > 0$ be two constants, and assume that*

$$k_{on_i} = k_{on}\mathbb{1}_{i<i_{max}} \quad \text{and} \quad k_{dep_i} = k_{dep}\mathbb{1}_{i\leq i_{max}}, \tag{7}$$

*If $i_{max} = \infty$, the equilibrium $\tilde{c} \in E^+$ is then given by :*

$$\begin{cases} \tilde{c}_i = c_1^s\left(\dfrac{\tilde{c}_1}{c_1^s}\right)^i & \text{when } 2 \leq i \leq \infty, \\ \tilde{c}_1(\rho) = c_1^s\left(1 + \dfrac{c_1^s - \sqrt{{c_1^s}^2 + 4\rho c_1^s}}{2\rho}\right), \end{cases}$$

*where* $c_1^s = \dfrac{k_{dep}}{k_{on}}$. *This formula yields a good approximation for* $i_{max} \gg 1$.

*2. If Assumption 7 is modified to exclude exchange between* $c_1$ *and* $c_2$, *i.e., if we have*

$$k_{on_i} = k_{on}\mathbb{1}_{2\leq i<i_{max}} \quad \textit{and} \quad k_{dep_i} = k_{dep}\mathbb{1}_{2<i\leq i_{max}}, \tag{8}$$

*then the equilibrium state becomes, for* $i_{max} = \infty$ :

$$\begin{cases} \tilde{c}_i = P_1\left(1-\dfrac{\tilde{c}_1}{c_1^s}\right)\left(\dfrac{\tilde{c}_1}{c_1^s}\right)^{i-2} & \textit{when } 2 \leq i \leq \infty, \\ \tilde{c}_1(\rho) = \dfrac{1}{2}\left(\rho - P_1 + c_1^s\right) - \sqrt{\left(\rho - P_1 - c_1^s\right)^2 + 4c_1^s P_1}, \end{cases}$$

*where* $c_1^s = \dfrac{k_{dep}}{k_{on}}$, *and* $P_1 = \sum_2^\infty c_i(t=0)$.

If $i_{max}$ is finite - as it is most probably the case in our application - these formula have to be modified by similar calculations as shown below (for the sake of simplicity we do not detail these cases here). The approximation will be on the order of $O\big((\frac{\tilde{c}_1}{c_1^s})^{i_{max}}\big)$.

**Proof:** 1. Under assumption (7), we have $Q_i := \left(\dfrac{\mathrm{k_{on}}}{\mathrm{k_{dep}}}\right)^{i-1}$ for $1 \leq i \leq i_{max}$ with $\tilde{\mathrm{c}}_i = Q_i \tilde{\mathrm{c}}_1^i$ for all $1 \leq i \leq i_{max}$ and

$$\sum_{i=1}^{i_{max}} i Q_i \tilde{c}_1^i = \rho. \tag{9}$$

For a matter of finite mass, if $i_{max} = \infty$ we find that $\tilde{\mathrm{c}}_1 < \mathrm{c}_1^s := \dfrac{\mathrm{k_{dep}}}{\mathrm{k_{on}}}$.

Setting $z := \dfrac{\tilde{\mathrm{c}}_1}{\mathrm{c}_1^s} < 1$, the mass conservation condition (9) yields (for $i_{max} = \infty$):

$$\rho = \mathrm{c}_1^s z \sum_{i=1}^{i_{max}} i z^{i-1} = \mathrm{c}_1^s \frac{z}{(1-z)^2} \quad \Rightarrow \quad z^2 - \frac{\mathrm{c}_1^s + 2\rho}{\rho} z + 1 = 0.$$

The determinant of this quadratic equation yields two possible values for $z$ :

$$\begin{cases} \Delta = \dfrac{\mathrm{c}_1^{s^2} + 4\rho \mathrm{c}_1^s}{\rho^2} > 0 \\ z = 1 + \dfrac{\mathrm{c}_1^s \pm \sqrt{\mathrm{c}_1^{s^2} + 4\rho \mathrm{c}_1^s}}{2\rho}. \end{cases}$$

Since $z < 1$, it can only take the value of the smaller root of the quadratic equation, and we finally obtain

$$z = 1 + \frac{\mathrm{c}_1^s - \sqrt{\mathrm{c}_1^{s^2} + 4\rho\mathrm{c}_1^s}}{2\rho} \quad \Rightarrow \tilde{\mathrm{c}}_1 = \mathrm{c}_1^s \left(1 + \frac{\mathrm{c}_1^s - \sqrt{\mathrm{c}_1^{s^2} + 4\rho\mathrm{c}_1^s}}{2\rho}\right).$$

2. Let us now assume (8). Biologically, this means that there is neither elongation for monomers nor depolymerization for fibrils of size $i_{\min} = 2$. Hence, the total number of fibrils of size $2 \leq i \leq i_{max}$ remains constant in time :

$$\mathrm{k}_{\mathrm{on}_1} = \mathrm{k}_{\mathrm{dep}_2} = 0 \;\Rightarrow\; J_1(\mathrm{c}) = 0 \qquad \Rightarrow \frac{d}{dt}\sum_{i=i_2}^{i_{max}} \mathrm{c}_i = 0, \;\; \forall\, t \in [0, \infty[\,.$$

The equilibrium solution is given by

$$\tilde{\mathrm{c}}_i = T_i \tilde{\mathrm{c}}_1^{i-2} \tilde{\mathrm{c}}_2, \quad 2 \leq i \leq i_{max} \tag{10}$$

$$\frac{d}{dt}\sum_{i=2}^{i_{max}} \tilde{\mathrm{c}}_i = 0, \tag{11}$$

$$\|\tilde{\mathrm{c}}\| = \rho, \tag{12}$$

where $T_i = \left(\dfrac{\mathrm{k}_{\mathrm{on}}}{\mathrm{k}_{\mathrm{dep}}}\right)^{i-2}$ for $2 \leq i \leq i_{max}$.
From (11), the total number conservation implies that, for $i_{max} = \infty$,

$$\tilde{\mathrm{c}}_2 = \frac{\mathrm{P}_1}{\sum_{i=2}^{\infty}\left(T_i \tilde{\mathrm{c}}_1^{i-2}\right)} = P_1\left(1 - \frac{\tilde{\mathrm{c}}_1}{\mathrm{c}_1^s}\right),$$

where $\mathrm{P}_1 := \sum_{i=2}^{\infty} \mathrm{c}_i\,(t=0)$. Here again, this implies that $\tilde{\mathrm{c}}_1 < \mathrm{c}_1^s := \dfrac{\mathrm{k}_{\mathrm{dep}}}{\mathrm{k}_{\mathrm{on}}}$.
On the other hand, we have the mass conservation equation which implies, by setting $z := \dfrac{\tilde{\mathrm{c}}_1}{\mathrm{c}_1^s}$:

$$\|\tilde{\mathrm{c}}\| = \tilde{\mathrm{c}}_1 + \sum_{i=2}^{\infty} i\mathrm{P}_1 z^{i-2}\,(1-z) = \mathrm{c}_1^s z + \mathrm{P}_1 \frac{2-z}{1-z}.$$

Hence, the total mass conservation (12) gives :

$$\mathrm{c}_1^s z^2 + \left((\mathrm{P}_1 - \rho) - \mathrm{c}_1^s\right) z + (\rho - 2\mathrm{P}_1) = 0.$$

The determinant of this quadratic equation gives

$$\begin{cases} \Delta = (\rho - \mathrm{P}_1 - \mathrm{c}_1^s)^2 + 4\mathrm{c}_1^s\mathrm{P}_1 > 0 \\ z = \dfrac{1}{2\mathrm{c}_1^s}\left(\rho - \mathrm{P}_1 + \mathrm{c}_1^s - \sqrt{\Delta}\right). \end{cases}$$

The choice of the smaller root is, here again, guided by the fact that $z < 1$ by definition. To prove this, we observe that

$$z_+ = \frac{1}{2\mathrm{c}_1^s}\left(\rho - \mathrm{P}_1 + \mathrm{c}_1^s + \sqrt{\Delta}\right) \geq \frac{1}{2\mathrm{c}_1^s}\left(\rho - \mathrm{P}_1 + \mathrm{c}_1^s + |\rho - \mathrm{P}_1 - \mathrm{c}_1^s|\right).$$

Let us first assume that $\mathrm{c}_1^s \geq \rho - \mathrm{P}_1$, then

$$z_+ \geq \frac{1}{2\mathrm{c}_1^s}\left(\rho - \mathrm{P}_1 + \mathrm{c}_1^s + (\mathrm{c}_1^s - \rho + \mathrm{P}_1)\right) \geq 1.$$

On the contrary, if $\mathrm{c}_1^s \leq \rho - \mathrm{P}_1$,

$$z_+ \geq \frac{1}{2\mathrm{c}_1^s}\left(\rho - \mathrm{P}_1 + \mathrm{c}_1^s + (-\mathrm{c}_1^s + \rho - \mathrm{P}_1)\right) = \frac{\rho - \mathrm{P}_1}{\mathrm{c}_1^s} \geq 1.$$

Independently of the shape of the initial condition, the equilibrium solution of the system (3) behaves as an exponential as time becomes large, see Figures 3 and 4. This supports the estimation for fibril formation developed in Section 2 (Figure 2).
As the size-distribution in Section 2 was found to follow an exponential law: $y = \lambda e^{-\lambda i}$ (Figure 2 ) the critical monomer concentration $\mathrm{c}_1^s$ can be estimated by an identification with the equilibrium formula of Proposition (4.1) from which we obtain

$$\begin{aligned} e^{-\lambda i} = \left(\frac{\tilde{\mathrm{c}}_1}{\mathrm{c}_1^s}\right)^i \Rightarrow e^{-\lambda} = \frac{\tilde{\mathrm{c}}_1}{\mathrm{c}_1^s}, \\ \Rightarrow e^{-\lambda} = 1 + \frac{\mathrm{c}_1^s + \sqrt{\mathrm{c}_1^{s^2} + 4\rho\mathrm{c}_1^s}}{2\rho}, \\ \Rightarrow \rho e^{\lambda}\left(e^{-\lambda} - 1\right)^2 = \mathrm{c}_1^s. \end{aligned}$$

Hence, the admissible critical monomer concentration that comes by this identification is

$$\mathrm{c}_1^s = \rho\left(e^{-\lambda} + e^{\lambda} - 2\right).$$

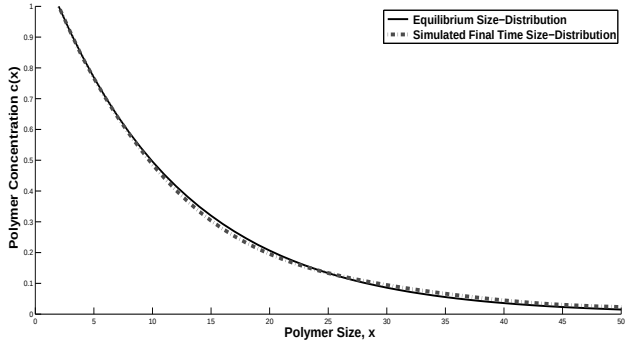


Figure 3: Equilibrium size-distribution of the polymers. In black full line the theorical equilibrium size-distribution, and in dashed-line the simulated size-distribution for large time ($t = 10^6 min$). Here $k_{dep}/k_{on} = 2.10^{-5}\mu M$, and $\rho = c_1(t = 0) = 22.10^{-6}\mu M$.

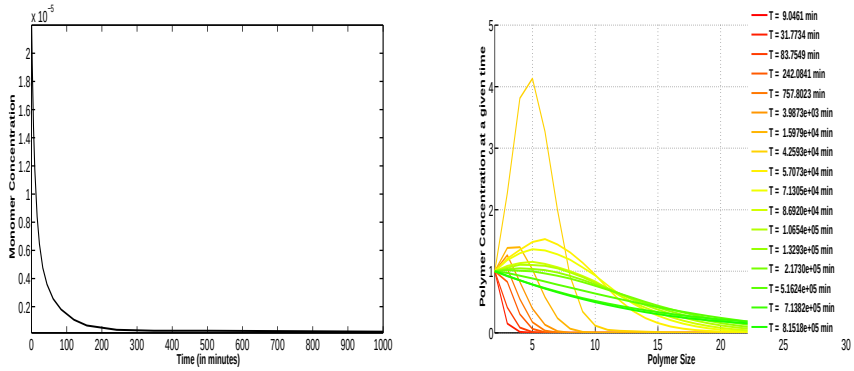


Figure 4: In the left panel, the time evolution of monomer concentration, with $k_{dep}/k_{on} = 2.10^{-5}\mu M$, and $\rho = c_1(t = 0) = 2.2.10^{-5}\mu M$. The curve is quickly decreasing, which is explained by a quick monomer consumption, to reach the critical monomer concentration $c_1^s = 2.10^{-5}\mu M$ first and then the equilibrium $\tilde{c}_1 = 1.5.10^{-7}\mu M$. In the right panel, the size-distribution of polymer concentration at different times: it reaches its equilibrium in much longer time than the concentration of monomers on the left.

Since $\lambda = \dfrac{1}{706.48}$, and $\rho = 2.2.10^{-5}\mu M$, we obtain

$$c_1^s = 4.41.10^{-11}\mu M.$$

Thus we obtain an estimation of both of the kinetic ratio and the equilibrium monomer concentration given by

$$\frac{\mathrm{k_{dep}}}{\mathrm{k_{on}}} = 4.41.10^{-11}\mu M,$$
$$\tilde{\mathrm{c}}_1 = 4.40.10^{-11}\mu M.$$

This implies that no monomers would have been detected experimentally. The numerical simulations using this experimental growth rate require a very large time to reach the equilibrium (see Figure 4). This is in agreement with the experimental observations, where from 4 up to 7 days are necessary to reach an apparent equilibrium in the polymerized mass (if we consider that the plateau in ThT fluorescence values can be an indication of a potential equilibrium for the polymerized mass).

## 5. Discussion

Due to the variety of the mechanisms involved in protein aggregation, different mathematical models have been developed. The main models and the main analysis of experimental data that have been tentatively applied to kinetics and/or size distribution of protein filament and fibril formation are outlined below. As amyloid filaments can be viewed as a linear shape formed by end-to-end association of units, we do not consider the case of protein aggregations that leads to globular unstructured clusters (such as the aspecific aggregation induced near the protein isoelectric point).

**Polymerization-depolymerization models.**
The growth of protein fibrils and the formation of synthetic polymers share analogous mechanisms which can embrace nucleation, polymerization, depolymerization, fragmentation and branching. Both protein fibrils and some synthetic polymers can be depicted as polymers called "living polymers", i.e., polymers that are not terminated and can freely polymerize or depolymerize. Starting from the theory of polymerization of synthetic compounds with their reactivity independent of the size of the polymers [20], Oosawa [34] mathematically described the size distribution of protein polymers and applied these ideas to experimental data on actin and flagellin polymerization. Using the law of mass action and thermodynamic laws on the interaction free energy of a monomer to a polymer, he obtained at equilibrium

$$\tilde{\mathrm{c}}_i = K_i \tilde{\mathrm{c}}_1^i, \quad \text{for} \quad i > i_0, \tag{13}$$

where $i_0$ is the size of the nucleus. Here $K$ is assumed to be independent of $i$ for sufficiently large sizes and is defined as $K = \exp\{\frac{\epsilon}{kT}\}$, with $-\epsilon$ being the interaction free energy of a monomer added to the end of a polymer of size $i-1$ to form a polymer of size $i$.

If we consider $K$ as equal to $\frac{\mathrm{k}_{\mathrm{dep}_j}}{\mathrm{k}_{\mathrm{on}_j}}$ for the size $j \geq 1$, this model of nucleation-polymerization-depolymerization is comparable to the formula of our mathematical model of polymerization-depolymerization (described in the previous section)

$$\tilde{\mathrm{c}}_i = \mathrm{c}_1^s \left(\frac{\tilde{\mathrm{c}}_1}{\mathrm{c}_1^s}\right)^i \quad \text{when } 2 \leq i \leq \infty \tag{14}$$

where $\mathrm{c}_1^s = \frac{\mathrm{k}_{\mathrm{dep}}}{\mathrm{k}_{\mathrm{on}}}$.

**Nucleation-polymerization-depolymerization vs. initiation-polymerization-depolymerization models.**

The first step of amyloid filament formation has not yet been clearly determined. Several studies propose, for prion conversion, the occurrence of a nucleation step, i.e., the formation of a nucleus, an oligomer of a critical size that permits a further elongation by addition of monomers (or small units), whereas oligomers of a size lower than the nucleus size would be instable. However the assumption of a nucleation step in amyloid aggregation is usually based on the presence of a lag phase. But a lag phase has been demonstrated to be possible even for models without nucleation and even when seed fibrils are already present [11].

As there is no proof for nucleation to be the step responsible for protein aggregation, other types of activation may be envisaged. Another type of activation step called initiation is well known in the field of chemical synthesis. Initiation can be seen as a kind of "ignition", due, for instance, to heating or to addition of an initiator, and generates a reactive intermediate. To compare the effect of nucleation and that of chemical initiation on aggregation kinetics and size distribution, Greer has compared a theoretical model of the nucleation-polymerization of the actin protein with a model of initiation-polymerization of the synthetic alpha-methylstyrene [26]. In both cases the models involve at least two reactions: i) a reversible activation step which is a nucleation or an initiation, and ii) a reversible polymerization step, called also elongation and propagation, by monomers. With a closed system, when the concentration of free monomers is considered to be capable of varying, once the elongation step has been entered, the molecular weight distribution of the polymers is assumed

to encounter three stages [31].
(i) As written by Brown and Szwarc for an "initiation-polymerization-depolymerization" system proceeding to the elongation step, the immediate but nonequilibrium result is the formation of a narrow Poisson distribution of polymers, if the number of monomers added by polymerization is very much higher than the ones removed by depolymerization [9]. At this first stage, we may define the generating function, $L(\lambda, \tau)$, of the size distribution by

$$L(\lambda, \tau) = \sum_{i=0}^{\infty} \lambda^i \mu_i(\tau),$$

where $\mu_i(\tau)$ is a dimensionless version of the concentration of polymers of size $i$ at time $\tau = \mathrm{k_{on}} M_0 t$. The authors show that this generating function closely follows the equations of irreversible polymerization, i.e., a system where depolymerization is not present [31]

$$L(\lambda, \tau) = B e^{-(1-\lambda)\upsilon(\tau)}.$$

Here

$$\upsilon(\tau) = \frac{\left(1 - e^{-B\tau}\right)}{B}$$

with $B = \dfrac{I_0}{M_0}$, $I_0$ being the initial initiator concentration and $M_0$, the initial monomer concentration. This is the generating function of a Poisson distribution of parameter $v(\tau)$.
(ii) Then possible sizes extend to a wider interval and free monomers reach an equilibrium concentration $M_{eq}$ given by

$$M_{eq} \cong \frac{\mathrm{k_{dep}}}{\mathrm{k_{on}}}$$

while number- and weight-average sizes remain near their Poisson values $1/B$ [31].
(iii) Over a longer time, the weight-average size (and the initiator concentration) evolve to a final equilibrium.
The times for the completion for the two first stages are of the order of $1/\left(\mathrm{k_{on}} I_0\right)$, while the time for the completion of the third stage is of the order of $1/\left(\mathrm{k_{dep}} B^2\right)$ [31]. More precisely, the times for completion for each stage are the following.

$$t_1 \cong ln \frac{\dfrac{B}{A}}{\mathrm{k_{on}} I_0} \quad , \qquad t_2 \cong ln \frac{\dfrac{1}{A}}{\mathrm{k_{on}} I_0} \quad , \qquad t_3 = \frac{1}{2\mathrm{k_{dep}} B^2},$$

with $A$ equal to $\frac{k_{dep}}{k_{on}M_0}$ [31]. The time for completion of the last step strongly depends on the ratio between the initiator concentration and the monomer concentration. In the case of amyloid proteins, we can take a value for $k_{dep}$ usually encountered in biology, for instance $10^{-4}\ sec^{-1}$. If we consider the monomers to be directly and fully converted into an activated conformer that can be depicted as a kind of initiator, the final equilibrium for the size distribution would be rapidly completed (within hours). On the contrary, if only a small percentage of the monomers is transformed in an activated form or if the rate of activation is slow, reaching the final size distribution can require years.

The final distribution of "initiation-polymerization-depolymerization" systems has been studied by both "mean field" and "non-mean-field" theories, as summarized by Greer [26]. Because both theories leads to a size distribution that may not easily be distinguished [26], we only discuss the result of the mean field theory. The mean field models (or self consistent field models) "treat the microscopic system as having an average intermolecular interaction, devoid of correlated interactions between or among molecules" [26] (as contrary to non-mean-field models). The mean field theory predicts that the equilibrium size distribution will be of the Flory-Schulz form

$$P_i = p^{i-1}\,(1-p)\,,$$

with $P_i$ the mole fraction of polymers of size i, and $p$ the probability that each monomer reacts to link to the growing polymer chains. When p tends to 1, the Flory-Schulz distribution becomes an exponential distribution [26].
An exponential final size distribution was experimentally observed both for actin and for alpha-methylstyrene, and a transition from Poisson to the exponential distribution was detected for actin [26]. However the rate-limiting step, i.e., the reaction step that slows down the whole aggregation process was found to be the activation step for actin and the elongation step for organic polymers. This difference may be explained by the fact that the activation step is a slow nucleation step for actin, whereas the activation step is an initiation step which immediately fixes the number of propagating species for alpha-methylstyrene [26]. As a conclusion, the final size distribution may have the same shape for protein polymerization that involves a nucleation as for synthetic polymerization of living polymers that starts by an initiation, but the kinetics induced by a nucleation may strongly differ from the one induced by an initiation. However in the cases of protein polymerization that are quickly activated, i.e., where nucleation does not exist (nucleus size considered as equal to 1) or is not a rate-limiting step [39], the kinetics might be similar to that of

synthetic living polymers.

**Nucleation-polymerization-depolymerization-coalescence model.**
We next focus exclusively on protein polymerization, describing the experimental results and analytical theory applied to them. Subsequent to the work of Oosawa on protein reversible polymerization, the aggregation of proteins has been tentatively modeled either by a model of polymerization and depolymerization, or by a polymerization and secondary nucleation pathways including fragmentation. The possibility of depolymerization unit by unit has been taken into account in modelling the polymerization of A$\beta$, the peptide involved in Alzheimers disease. The authors of [35] postulate, as other authors have done also for A$\beta$, that the experimental conditions (highly concentrated urea) used to launch fibril growth by unfolding the monomeric peptide induce an irreversible modification that leads a part of the peptides to an amyloid status and another part to a modified monomer or dimer status unable to adopt an amyloid status. In order to model the concentration in monomers, dimers and polymers that has been deduced from experimental data at three different initial monomer concentrations, the authors tested the following five-step model:
(i) rapid commitment to a stable monomer or dimer or an unstable intermediate (as a kind of chemical initiation step),
(ii) cooperative association of intermediate into a multimeric "nucleus",
(iii) reversible elongation by addition of intermediates to form filaments,
(iv) lateral aggregation of filaments into fibrils,
and (v) fibril elongation via end-to-end association [35].

Model parameters were derived by fitting experimental data to the model equation, using a parameter estimation package. First, the "refolding" parameters involved in step (i) were determined. Then, as lateral aggregation and end-to-end association seemed experimentally negligible at the lowest protein concentration, the parameters involved in filament initiation and elongation could be obtained from the data measured at the lowest protein concentration. Lastly, the parameters describing lateral aggregation and end-to-end association were determined, after considering that filament initiation and filament elongation to be essentially irreversible, i.e., considering depolymerization as negligible. Their model was able to reproduce their experimental data as, at short times, the average length of filaments or fibrils formed at the lowest monomeric concentration is much higher than the one at higher concentrations, due to the competition between nucleus formation and filament elongation for the consumption of intermediates. One can remark that also for our simple

"polymerization-depolymerization" model, at equilibrium, the smaller the mass $\rho$, the longer the average size i, when considering fixed $c_1^s$ (see Proposition 4.1). In the case of the model that contains also nucleation and coalescence, at long times, the average size is not necessarily higher for the lowest monomeric concentration [35]. This underlines that, depending on the initial concentration of monomers, the relative importance of elongation, or polymerization, compared to the one of initiation and lateral filament aggregation, or coalescence, may greatly vary, regarding fibril lengths [35].

**Experimental data analysis by Xue and coworkers [50] and comparison to our data and our "polymerization-depolymerization" model.** Xue and co-workers [50] have analysed by statistical tests the fibril size distribution obtained after an extensive experimental study.The sizes of $\beta$-2-microglobulin fibrils formed in vitro by adding fibril seeds to monomers have been experimentally measured, after a certain time under quiescent conditions or agitation. The authors have developed and applied a correction factor on their data to correct for the fact that longer fibrils were underestimated compared to smaller fibrils. This experimental bias is due to two reasons: longer fibrils may less efficiently deposit on the mica surface of the AFM (atomic force microscope) and longer fibrils have a higher probability to not be successfully traced on images because of fibril overlap and cut-off by image boundaries. This bias can also be encountered for the technique used for our experiments described in Section 2 above, TEM (transmission electron microscopy). To avoid fibril overlap and cut-off by image boundaries, a low fibril concentration was used and our images were carefully taken to avoid as much as possible cut-off of long fibrils on image sides. Concerning difference in fibril deposition on the carbon-formvar surface as a function of fibril size, this problem is particularly present for samples containing a high concentration of fibrils and especially a high concentration of monomers as the monomers strongly compete with longer fibrils for adsorption on the carbon-formvar surface [33]. We have not corrected for a bias that would be due to difference in fibril deposition as a function of fibril length, but we have minimized this problem by using a low fibril concentration on the grids (0.1 $\mu$M). Xue and coworkers compared the overall distribution shape to distribution models that belongs or can belong (by fixing some parameters) to the exponential family [50]. The Weibull distribution and, to a lesser extent, the Gamma distribution, consistently fitted the data, whereas the other models, the normal, log-normal, exponential and Rayleigh distributions, fitted significantly more poorly the measured sample distributions as judged by the Kolmogorov-Smirnov test. Note that our experimental data also may be

viewed as fitting a Weibull distribution, since the Weibull distribution, defined by the formula below, with a shape parameter $k$ equal to 1 is an exponential:

$$f\left(i;\lambda,k\right)=\frac{k}{\lambda}\left(\frac{i}{\lambda}\right)^{k-1}exp^{-(i/\lambda)^{k}}$$

where $i \geq 0$ and $\lambda > 0$. Because our data are truncated for small sizes (due to experimental limit detection), it is also possible that our data might fit a Weibull distribution with various values of k.

The dome-shaped experimental distributions observed by Xue and coworkers are different from the pure exponential shape of our mathematical polymerization-depolymerization model at equilibrium. At least two hypotheses can easily explain this difference: i) their system also follows a polymerization-depolymerization model however their experimental state has not reached equilibrium, or ii) an additional mechanism needs to be added to our polymerization-depolymerization model. Regarding the first hypothesis, one can envisage that the distribution experimentally observed by the authors is an intermediate step before reaching a pure exponential distribution at equilibrium (as there is no clue to the stage reached by the samples). Concerning the second hypothesis, an additional mechanism, fragmentation, has been shown to play a major role in kinetic of fibrils that grow under agitation, as explained below.

Because the length distribution of the fibrils obtained by addition of a small percentage of seeds (0.1%) to monomers has approximately the same shape as the one obtained by addition of a high percentage of seeds (10%) to monomers, Xue and coworkers proposed that fragmentation might be a significant event even when samples are not agitated [50]. However the average size between these two samples may seem different. It would be interesting to compare these data to a sample that does not contain added seeds and has grown under quiescent conditions. That would help to estimate to what extent the number of fibrils that are due to seed elongation contributes to the number of total fibrils (fibrils due to seed elongation and fibrils directly formed from monomers) and to the incorporation of monomers. This would help to distinguish between the effect of seed elongation, the effect of elongation of new nucleus formed from monomers and the effect of a hypothetical fragmentation under quiescent conditions.

When fibrils grown under quiescent conditions are subsequently mechanically agitated, the size distribution shifted to lower sizes due to fibril fragmentation [49, 51]. Also in the case of fibril growth under quiescent conditions, it has been proposed, to explain the kinetics of experimentally measured average fibril length, that fragmentation can occur [44]. However the relative importance of fragmentation compared to depolymerization is unknown.

**Nucleation-polymerization-fragmentation vs. nucleation-polymerization-depolymerization.**
To compare the theoretical effect of fragmentation with the one of depolymerization, Cohen and co-workers [11] have constructed several models. Their models consist in nucleation, polymerization, in the absence or the presence of fragmentation and/or depolymerization. The depolymerization parameter was not included in the expression of the concentration of fibrils, $P$, assuming that once formed, a fibril cannot vanish by the effect of depolymerization. The authors of [11] demonstrated that, in the absence of fragmentation, the following equation which shares some analogy with Johnson-Melh-Avrami-Kolmogorov equation for crystallisation, is obtained

$$M(t) = m_{tot}\left(1 - exp\left(-\mathrm{k_{on}} k_n m_{tot}^{n_c} t^2 - 2\mathrm{k_{on}} P(0) t - \frac{M(0)}{m_{tot}}\right)\right). \qquad (15)$$

Here $M(t)$ is the mass concentration of polymers, $P(t)$ is the number concentration of polymers, $\mathrm{k_{on}}$ is the polymerization coefficient, $k_n$ is the coefficient for nucleation, $m_{tot}$ is the total monomer concentration (free monomers plus monomers included in polymers), and $n_c$ is the critical nucleus size.
If total monomer concentration is kept constant, through, for example, protein synthesis or experimental supply to the system, or in the early time limit, fibril growth tends to a polynomial form in time given by

$$M_0(t) = M(0) + \mathrm{k_{on}} k_n m_{tot}^{n_c+1} t^2 + 2\mathrm{k_{on}} m_{tot} P(0) t + n_c k_n m_{tot}^{n_c} t. \qquad (16)$$

Because filament growth can be bidirectional with a varying degree of reactivity depending on which of the two extremities is considered (as shown by microscopy on yeast prion fibers [27, 14]), Cohen and co-workers included a factor 2 in the polymerization and depolymerization expressions. In the mass equation of Cohen and co-workers, the $t^2$ dependence of the Oosawa solution equation is recovered even though Oosawa considered one unique reactive site per polymer.

The width of the filament length distribution is represented by the variance $\sigma^2(t)$ of the filament population. In the presence of fragmentation and by considering the effect of fragmentation as limited (assuming $m(0)\mathrm{k_{on}} \gg k_{fragm}$), and the monomer concentration as approximately constant, the following equation was obtained

$$\sigma(t) = \frac{1}{\sqrt{3}}\mu(t), \tag{17}$$

where $\mu(t)$ is the mean length of the distribution. This equation means that, for this system of fragmenting filaments growing in constant free monomer concentration, the ratio of the mean filament length to the standard deviation of the distribution of filament lenghts is constant. At longer times, when monomers become depleted, both mean filament length and the standard deviation of the filament length distribution decrease because of fragmentation [11].

As discussed in [12], for systems with a constant concentration of total protein (free monomers plus monomeric units included in polymers) that encompass nucleation, polymerization and fragmentation, the concentration of polymers of each size can be written at steady state $t = \infty$ as

$$\frac{\mathrm{c}_i}{\mathrm{P}} = \frac{n_c \left((n_c - 1)\, n_c\right)^{i-n_c} \left(n_c - n_c^2 + i + i^2\right) \left(n_c^2 - 1\right)!}{(1 + (n_c - 1)\, n_c + i)!} \tag{18}$$

with $\mathrm{c}_i$, the concentration of polymers of size $i$, P the total concentration of polymers and $n_c$, the critical nucleus size. As can be seen, the only reaction that controls the shape of the length distribution at steady state is the nucleation reaction. The polymerization and the fragmentation only affect the time-scale necessary to reach the steady state length distribution.
In the continuum limit, after simplifications [12], the length distribution at steady state in such system which includes nucleation, polymerization and fragmentation, has a biased Gaussian form given by

$$\frac{\mathrm{c}_i}{\mathrm{P}} = \frac{e^{\frac{2\left(-i^2 + n_c^2\right)}{(1-2n_c)^2}} \left(4i^2 - (1-2n_c)^2\right)}{(1-2n_c)^2\, n_c}. \tag{19}$$

The fibril size distribution should therefore exhibit a Gaussian form when fragmentation is involved and, as written above, an exponential form in the absence of fragmentation. This difference could have been easily expected as fragmentation induces the formation of small filaments at the expense of long filaments. For our experimental study, due to the data truncation for concentrations of small polymers, we cannot confirm or disaffirm an occurrence of fragmentation. However, as the fibril formation has been performed under a relatively smooth agitation, fragmentation may be limited.

**Additional process: secondary nucleation.**
The effect of an hypothetical secondary process that creates new structures of a critical size $n_2$ through "nucleation" on the surface of existing polymers was also discussed by Cohen and coworkers [11, 12]. If $n_2 \approx n_c$, the distribution is expected to be of qualitatively similar form, with an equilibrium distribution of the exponential type as in the case of a system with primary nucleation [11, 12].

**Conclusion.**
A number of mathematical models have been developed in the recent years and have been applied with a, at least partial, success to experimental data on fibril distribution. According to these models, an exponential distribution is expected in the case of polymerization-depolymerization reactions, whether or not nucleation is present. If fragmentation is not negligible, fibril sizes should then exhibit a peaked shape. An exponential distribution and therefore a polymerization-depolymerization model were found to apply to actin and flagellin. It may also apply to our data on prion fibril formation. Though, due the truncation in our data, it might be possible that other models, such as one including a fragmentation step, may also reproduce the prion data.

Generally speaking the relative importance of each of the possible amyloid reactions remains to be clarified. To reach this goal, mathematical models should be tested on experimental size distributions obtained ideally at various times and various monomer concentrations. Because biochemical and biophysical studies usually require and therefore use amyloidic protein concentrations much higher than the ones encountered in vivo, these techniques can offer a way to identify the existing amyloid reactions, but deducing the relative importance of each amyloid reaction from these techniques may not be relevant to in vivo conditions. The development of high-resolution techniques on cells combined to automated data analysis software could offer numerous data on fibril size distribution for in-cellula environments. These technical developments would help to identify the characteristics of the most reactive amyloid forms that recruit native monomers within one cell. To extrapolate to in vivo conditions, it will still remain to determine whether the most reactive amyloid species within one cell is also responsible for the infectious character within one individual and whether it is the species able to spread between individuals.

**Acknowledgments**

The authors thank Joan Torrent Y Mas (INSERM-INRA), for the help on fibril production and for discussion. This research was supported by the ERC Starting Grant SKIPPER$^{AD}$ (fully for M. Doumic and S. Prigent and partly for W. Haffaf and H. Rezaei).

**References**

[1] T. Alper, W. A. Cramp, D. A. Haig, and M. C. Clarke. Does the agent of scrapie replicate without nucleic acid ? *Nature*, 214:764–766, 1967.

[2] M. Anderson, O. V. Bocharova, N. Makarava, L. Breydo, V. V. Salnikov, and I. V. Baskakov. Polymorphism and ultrastructural organization of prion protein amyloid fibrils: an insight from high resolution atomic force microscopy. *Journal of Molecular Biology*, 358(2):580–96, 2006.

[3] O. Andreoletti, L. Litaise, H. , Simmons, F. Corbiere, S. Lugan, F. Costes, F. Schelcher, D. Vilette, J. Grassi, and C. Lacroux. Highly efficient prion transmission by blood transfusion. *PLoS Path*, 8:e1002782, 2012.

[4] D. Balagué, J. Cañizo, and P. Gabriel. Fine asymptotics of profiles and relaxation to equilibrium for growth-fragmentation equations with variable drift rates. *Kinetic Related Models*, 2013. in press.

[5] J. M. Ball and J. Carr. Asymptotic behaviour of solutions to the becker-dring equations for arbitrary initial data. *Proc. Roy. Soc. Edinburgh Sect.*, A 108:109–116, 1988.

[6] J. M. Ball, J. Carr, and O. Penrose. The Becker-Dring cluster equations: Basic properties and asymptotic behaviour of solutions. *Commun. Math. Phys.*, 104:657–692, 1986.

[7] R. Becker and W. Dring. Kinetische Behandlung der Keimbildung in berst-tigten Dmpfern. *Ann. Phys.*, 24:719–752, 1935.

[8] V. Beringue, L. Herzog, E. Jaumain, F. Reine, P. Sibille, A. Le Dur, J.-L. Vilotte, and Laude H. Facilitated cross-species transmission of prions in extraneural tissue. *Science*, 335:472–475, 2012.

[9] W.-B Brown and M. Szwarc. Molecular weight distribution of "living" polymers. *Trans. Faraday Soc.*, 54:416–419, 1958.

[10] J J Burton. *Nucleation theory*. Statistical mechanics, Part A : Equilibrium techniques. Springer US, 1977.

[11] S. Cohen, M. Vendruscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, and T. P. J. Knowles. Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments. *The Journal of Chemical Physics*, 135(6):065105, 2011.

[12] S. I. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles. Nucleated polymerization with secondary pathways. III. Equilibrium behavior and oligomer populations. *The Journal of Chemical Physics*, 135(6):065107, 2011.

[13] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman. Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLOS Biology*, 2:e321, 2004.

[14] A. H. DePace and J. S. Weissman. Origins and kinetic consequences of diversity in Sup35 yeast prion fibers. *Nature Structural Biology*, 9(5):389–96, 2002.

[15] M. Doumic and P. Gabriel. Eigenelements of a general aggregation-fragmentation model. *Mathematical Models and Methods in Applied Sciences*, 20(05):757, 2009.

[16] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, under revision, 2014.

[17] M. Doumic, M. Hoffmann, P. Reynaud, and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *SIAM J. on Numer. Anal.*, 50(2):925–950, 2012.

[18] M. Doumic, B. Perthame, and J.P. Zubelli. Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems*, 25(4):045008, 2009.

[19] D. El Moustaine, V. Perrier, I. Acquatella-Tran van Va, F. Meersman, V. G. Ostapchenko, I. V. Baskakov, R. Lange, and Torrent J. Amyloid features and neuronal toxicity of mature prion fibrils are highly sensitive to high pressure. *Journal of Biological Chemistry*, 286:13448–13459, 2011.

[20] P. J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, 1953.

[21] Y. Furukawa and N. Nukina. Functional diversity of protein fibrillar aggregates from physiology to rna granules to neurodegenerative diseases. *Biochimica et Biophysica Acta*, 1832:1271–1278, 2013.

[22] N. Gill, Y. Spence, A. Richard-Loendt, A. Kelly, R. Dabaghian, L. Boyes, J. Linehan, M. Simmons, P. Webb, P. Bellerby, N. Andrews, D. A. Hilton, J. W. Ironside, J. Beck, M. Poulter, S. Mead, and S. Brandner. Prevalent abnormal prion protein in human appendixes after bovine spongiform encephalopathy epizootic: large scale survey. *BMJ*, 347:f5675, 2013.

[23] J. E. Gillam and C. E. MacPhee. Modelling amyloid fibril formation kinetics: mechanisms of nucleation and growth. *J. Phys. Condens. Matter*, 2013:25, 373101.

[24] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39:1608–1632, 2011.

[25] A. Goldenshluger and O. Lepski. Uniform bounds for norms of sums of independent random functions. *Ann. Probab.*, 39:2318–2384, 2011.

[26] S. C. Greer. Reversible polymerizations and aggregations. *Annual Review of Physical Chemistry*, 53:173–200, 2002.

[27] Y. Inoue, A. Kishimoto, J. Hirao, M. Yoshida, and H. Taguchi. Strong growth polarity of yeast prion fiber revealed by single fiber imaging. *The Journal of Biological Chemistry*, 276(38):35227–30, 2001.

[28] I. Koch and W. Reisig. *Modeling in Systems Biology - The Petri Net Approach*. Springer, 2011.

[29] T. Liu and Bitan G. Modulating self-assembly of amyloidogenic proteins as a therapeutic approach for neurodegenerative diseases: Strategies and mechanisms. *Chem Med Chem*, 7:359–374, 2012.

[30] F. Meersman and C. M. Dobson. Probing the pressure-temperature stability of amyloid fibrils provides new insights into their molecular properties. *Biochimica et Biophysica Acta*, 1764:452–460, 2006.

[31] A. Miyake and W. H. Stockmayer. Theoretical reaction kinetics of reversible living polymerization. *Die Makromolekulare Chemie*, 88:90–116, 1965.

[32] A. S. Mostaert, C. Giordani, R. Crockett, U. Karsten, R. Schumann, and S. P. Parvis. Characterisation of amyloid nanostructures in the natural adhesive of unicellular subaerial algae. *Journal of Adhesion*, 85:465–485, 2009.

[33] M. Necula and J. Kuret. Electron microscopy as a quantitative method for investigating tau fibrillization. *Analytical Biochemistry*, 329(2):238–46, 2004.

[34] F. Oosawa. Size distribution of protein polymers. *Journal of Theoretical Biology*, 27(1):69–86, 1970.

[35] M. M. Pallitto and R. M. Murphy. A mathematical model of the kinetics of beta-amyloid fibril growth from the denatured state. *Biophysical Journal*, 81(3):1805–22, 2001.

[36] O. Penrose. The Becker-Dring equations for the kinetics of phase transitions. *Math. Proc. Camb. Phil. Soc.*, 96, 2001.

[37] O. Penrose and J.L. Lebowitz. *Towards a Rigorous Molecular Theory of Metastability*. Gauthier-Villars, 1979.

[38] A. T. Petkova, R. D. Leapman, Z. Guo, W. M. Yau, M. P. Mattson, and R. Tycko. Self-propagating, molecular-level polymorphism in alzheimerss beta-amyloid fibrils. *Science*, 307:262–265, 2005.

[39] S. Prigent, A. Ballesta, F. Charles, N. Lenuzza, P. Gabriel, L.-M. Tine, H. Rezaei, and M. Doumic. An efficient kinetic model for assemblies of amyloid fibrils and its application to polyglutamine aggregation. *Plos One*, 7:0043273, 2012.

[40] H. Rezaei, D. Marc, Y. Choiset, M. Takahashi, G. Hui Bon Hoa, T . Haertle, J. Grosclaude, and P. Debey. High yield purification and physicochemical properties of full-length recombinant allelic variants of sheep prion protein linked to scrapie susceptibility. *European Journal of Biochemistry*, 267:2833–2839, 2000.

[41] K. Schwartz and B. R. Boles. Microbial amyloids functions and interactions within the host. *Current Opinion in Microbiology*, 2012:93–99, 2012.

[42] B Silverman. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1986.

[43] J. D. Sipe and A. Cohen. Review: History of the amyloid fibril. *Journal of Structural Biology*, 130:88–98, 2000.

[44] J. F. Smith, T. P. J. Knowles, C. M. Dobson, C. E. Macphee, and M. E. Welland. Characterization of the nanoscale properties of individual amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America*, 103(43):15806–11, 2006.

[45] Y. Sun, N. Makarava, C.I. Lee, P. Laksanalamai, F. T. Robb, and I. V. Baskakov. Conformational stability of prp amyloid fibrils controls their smallest possible fragment size. *J Mol Biol.*, 376:11551167, 2008.

[46] K. K. M. Sweers, M. L. Bennink, and V. Subramaniam. Nanomechanical properties of single amyloid fibrils. *J. Phys.: Condens. Matter*, 24:243101, 2012.

[47] National CJD Surveillance Unit. http://www.cjd.ed.ac.uk/. 2013.

[48] L. Wasserman. All of statistics : A concise course in statistical inference brief contents. *Simulation*, C(1):461, 2004.

[49] W.-F. Xue, S. W. Homans, and S. E. Radford. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 105(26):8926–31, 2008.

[50] W.-F. Xue, S. W. Homans, and S. E. Radford. Amyloid fibril length distribution quantified by atomic force microscopy single-particle image analysis. *Protein Engineering, Design & Selection : PEDS*, 22(8):489–96, 2009.

[51] W.-F. Xue and S. E. Radford. An imaging and systems modeling approach to fibril breakage enables prediction of amyloid behavior. *Biophysical Journal*, 105(12):2811–9, 2013.