An Efficient Kinetic Model for Assemblies of Amyloid Fibrils and Its Application to Polyglutamine Aggregation

Stéphanie Prigent¹, Annabelle Ballesta², Frédérique Charles³, Natacha Lenuzza⁴, Pierre Gabriel⁵, Léon Matar Tine⁶, Human Rezaei¹, Marie Doumic²*

1 Institut National de Recherche Agronomique, Jouy-en-Josas, France, 2 Institut National de Recherche en Informatique et Automatique, Rocquencourt, France, 3 Université Pierre et Marie Curie, Paris, France, 4 Commissariat à l’Énergie Atomique, Gif-sur-Yvette, France, 5 Institut National de Recherche en Informatique et Automatique Rhônes-Alpes, Lyon, France, 6 Université Gaston Berger, Saint-Louis, Sénégal

Abstract

Protein polymerization consists in the aggregation of single monomers into polymers that may fragment. Fibrils assembly is a key process in amyloid diseases. Up to now, protein aggregation was commonly mathematically simulated by a polymer size-structured ordinary differential equations (ODE) system, which is infinite by definition and therefore leads to high computational costs. Moreover, this Ordinary Differential Equation-based modeling approach implies biological assumptions that may be difficult to justify in the general case. For example, whereas several ordinary differential equation models use the assumption that polymerization would occur at a constant rate independently of polymer size, it cannot be applied to certain protein aggregation mechanisms. Here, we propose a novel and efficient analytical method, capable of modelling and simulating amyloid aggregation processes. This alternative approach consists of an integro-Partial Differential Equation (PDE) model of coalescence-fragmentation type that was mathematically derived from the infinite differential system by asymptotic analysis. To illustrate the efficiency of our approach, we applied it to aggregation experiments on polyglutamine polymers that are involved in Huntington’s disease. Our model demonstrates the existence of a monomeric structural intermediate acting as a nucleus and deriving from a non polymerizing monomer. Furthermore, we compared our model to previously published works carried out in different contexts and proved its accuracy to describe other amyloid aggregation processes.

Introduction

Protein aggregation and misfolding are involved in several fatal human disorders, such as Alzheimer’s, Prion, Huntington’s diseases [1,2]. Certain types of aggregates display specific structural traits (e.g. a β-sheet enriched secondary structure) that characterize amyloid assemblies. Recent progress in solid state Nuclear Magnetic Resonance (NMR) has led to a better understanding of amyloid assemblies at the molecular level [3]. However, this structural knowledge constitutes only a snapshot of the dynamic processes. Protein aggregation involves a large amount of chain reactions, e.g. conformational exchange, nucleation (which is the formation of a first stable intermediate), polymerization by monomer, dimer or monomer addition, coalescence, depolymerization (by loss of mono, di or oligomers), fragmentation (breakage into two or more polymers), protein degradation.

To explore the dynamics of amyloid assemblies, nucleation/polymerization reaction schemes have been applied, and to model them, ordinary differential equations (ODEs) have been used for many years [4]. An ODE means an equation containing only one independent variable (e.g. the chemical concentration of molecules) and its derivatives. Therefore in the case of polymerization, the number of equations should be at least equal to the number of sub-units constituting the longest polymer. This value is extremely large in the case of amyloid fibrils (amyloid fibril sizes can go up to 1 μm length [5]), therefore simplifying assumptions are commonly admitted, e.g. constant reaction rates, meaning that polymers of any size behave roughly in the same way [6–9]. Although such assumptions allow the system to be reduced from an infinite set of ODEs to a couple of equations [4,7], assumptions of this nature are difficult to justify biochemically.

We propose here a new and global framework that can be adapted to most protein polymerization reactions. This method relies on partial differential equations (PDEs). In contrast to an ODE, a PDE permits formulation of problems involving functions of several variables. Instead of handling an infinite set of ODEs, we rely on partial differential equations (PDEs). The size variable of fibers replaces the infinite number of ODEs. To derive our model, we tune asymptotic methods from previously published works [10,11]. A fully general model, which is much easier to handle both theoretically and numerically, is obtained. It allows much faster computations than for the full ODE set of equations. Moreover, recent analytical tools developed for PDE analysis can be applied.


Editor: Yoshitaka Nagai, National Center of Neurology and Psychiatry, Japan

Received May 8, 2012; Accepted July 18, 2012; Published November 13, 2012

Copyright: © 2012 Prigent et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The work was partially financed by the French Agence Nationale de la Recherche, ANR grant 09-BLAN-0218-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: marie.doumic@inria.fr
The obtention of size-distributions of polymers is a fundamental step [12], as it makes it possible to estimate quantitative reaction rates and build a predictive model by the means of recently developed inverse problem techniques [13].

To illustrate our method, we first formally derive the PDE model in a general case, and then apply our method to expanded polyglutamine (PolyQ) diseases. Finally, we compare our results to existing work [7,8].

Results

The Infinite ODE System

Let us first recall how one can write the differential system describing all the reactions that occur during nucleated protein polymerization. We denote $c_i$ the protein monomeric concentration and $c_{z_i}$ the one of a misfolded monomeric species which displays the ability to polymerize. $c_i$ monomers transform into this monomeric species $c_i^*$ at the rate $k_{li}^+$, and $c_{z_i}$ transform back to $c_i$ at the rate $k_{li}^-$. 

$c_i$ represents the concentration of polymers made up of $i$ monomers. We assume that polymers and monomers are degraded with a size-dependent degradation rate denoted $k_{m}$. The misfolded monomers $c_{z_i}$ are able to polymerize to give rise to a nucleus $c_{n}$, composed of $i_0$ monomeric units, with the rate $k_{N}^i$. As proposed by Oosawa and co-authors [4], a nucleus is generated by the addition of an object to highly unstable entities that are too transitory to be observed. The object stabilizing the highly unstable entities can be a monomer ($c_1$). If we consider a nucleus $c_{n}$ with a size $i_0$, its formation does not consist in a sequential addition of $c_1$ till $c_{n}$; (where it would be represented by $c_1 \rightarrow c_2 \rightarrow c_3 \rightarrow \cdots \rightarrow c_{i_0}$), but follows an $i_0$ order kinetic (where $i_0 c_1 \rightarrow c_{i_0}$).

This nucleus can dissociate at the rate $k_{N}^i$. We make the reasonable assumption that there is an equilibrium between monomers and oligomers [4].

$$
\frac{d c_i}{d t} = \frac{k_{li}^+}{k_{li}^-} c_{i_0} c_{i_0}^{*} + \cdots + \frac{k_{N}^i}{k_{N}^i} c_{i_0}^{*} c_{i_0}
$$

(1)

Polymers of size $i$ larger than $i_0$ can polymerize or depolymerize, which is the gain or the loss of a single monomeric unit: the elongating species is assumed here to be $c_{i_0}$ (our model is easy to adapt to other cases, e.g. if the elongating species is a dimer or an oligomer [8]). Those reactions occur at the rate $k_{on}^i$ and $k_{dep}^i$ respectively.

$$
\frac{d c_i}{d t} = \frac{k_{on}^i}{k_{dep}^i} c_{i} c_{i+1} + \frac{k_{on}^{i+1}}{k_{dep}^{i+1}} c_{i+1} c_{i+1}
$$

(2)

Polymers can also coalesce with other polymers or break into two smaller polymers. For the sake of simplicity, we assume that a polymer can only break into two pieces at the exact same time (a breakage into 3 or more pieces is generally much more hazardous, so that it can be neglected). Coagulation of two polymers of respective size $i$ and $j$ occurs at the rate $k_{col}^{i,j}$. Fragmentation of a polymer of size $i$ gives rise to smaller polymers of size $j$ and $i-j$ (where $2 \leq j \leq i_0$), at the rate $k_{off}^{i,j}$.

$$
\frac{d c_i}{d t} = c_i \frac{k_{on}^{i}}{k_{dep}^{i}} c_{i+1} + c_j \frac{k_{on}^{j}}{k_{dep}^{j}} c_{i+j}
$$

(3)

We could have kept the same notation for fragmentation and depolymerization, by denoting $k_{off}^{i,j} = k_{off}^{i-1,j} = \frac{k_{dep}}{2}$. We preferred however to distinguish them, because they involve reactions of different kinds, so that the orders of magnitude may appear different.

Let us define $k_{off}^{i,j} = \sum_{i=1}^{j} k_{off}^{i,j}$. This represents the total rate with which a polymer of size $j$ can break to give smaller polymers. By symmetry we have that $k_{off}^{i,j} = k_{off}^{i,j}$ and $k_{off}^{i,j} = k_{off}^{i,j}$.

The following model is the exact deterministic transcription of the previously considered reactions. It could be completed by other reactions (polymerization pathways, other types of conformational exchange, for instance) to adapt to any possible case. The variation $\frac{d c_i}{d t}$ of the species $c_i$ (or $c_{i_0}, c_{i_0}^{*}$) depends on two phenomena: 1) their rates of consumption, including depolymerization into a smaller polymer (or transformation into $c_{i_0}$ in the case of $c_{i_0}^{*}$), polymerization into a higher polymer (or transformation into $c_{i_0}$ in the case of $c_{i_0}^{*}$) and degradation $k_{m}$, and 2) their rates of production, i.e. polymerization from smaller polymer (or transformation from $c_{i_0}$ in the case of $c_{i_0}^{*}$) and depolymerization from higher polymer (or transformation from $c_{i_0}$ in the case of $c_{i_0}^{*}$). This induces the following equations.

$$
\frac{d c_i}{d t} = -k_{li}^+ c_{i+1} + k_{li}^- c_i + k_{m}^i c_i

$$

(4)

$$
\frac{d c_i}{d t} = k_{on}^i c_i - k_{N}^i c_i - i_0 k_{on}^{i_0} c_{i_0} + \sum_{j=0}^{\infty} i_{j} k_{off}^{i,j} c_j

$$

(5)

$$
\frac{d c_{i_0}}{d t} = k_{off}^{i_0,j_0} c_{i_0}^{j_0} - k_{on}^{i_0} c_{i_0}^{i_0} + \sum_{j=0}^{\infty} k_{off}^{i_0,j} c_{i_0}^{j_0} c_j

$$

(6)
can be derived from the infinite set of ODEs if the two following
have defined the small parameter
of a polymer by a continuous variable
for
such that
Cst
polymers -

natural it is to view the size of polymers as a continuous quantity.

First, for most polymer sizes
, there is only a slight difference
between what happens for
mers and for
-mers. In other
terms, even if quantities and reaction rates vary, it occurs in a
“continuous” manner, implying only slight changes from one size
i
to its neighbor sizes
and
except for a small number of values. For instance, for
garation coefficients
, it is formalized as: There exists a constant, denoted below
Cst>0,
such that

For all
, 

This assertion allows a continuous viewpoint on the equations
for
. It also means that disruptions in the concentrations or in
the coefficients can only appear at some specific points, that will have
to be identified, and that are meaningful biologically. Though, this
assertion appears to be natural since the conformational changes
in polymers only occur at specific sizes [16]. Moreover, having a
look at experimental size distributions (Figure 1) confirms how
natural it is to view the size of polymers as a continuous quantity.

The second and quite standard assumption is that at the
beginning of the reaction, when polymer concentrations remain
small compared to monomers, polymerization is the main process,
whereas fragmentation and coalescence are secondary processes
[4,6]. This assumption can be replaced if necessary by a similar
one, such as the existence of a dominant polymerization by
mer addition, with
a relatively small oligomer. In such a case, the
polymerization terms
would be replaced by
in the
equations, and a similar treatment can apply.

We refer to supplementary data S1 for a rigorous mathematical
formulation of these two assertions. They are obtained when the
system of equations is rescaled, and this allows us to estimate the
relative contribution of each process to the overall dynamics.

Let us turn to the nucleus
. In this equation, the two assertions
make it possible to ignore the influence of fragmentation and
coalescence. Then as we are in the case where
, the
time-dependency of the equation for
is much faster than the one for
: it can be written

Hence, it is valid to suppose that it reaches its equilibrium
instantaneously, and we can replace Equation (6) by

We thus obtain the following equality, which generalizes well-
established formulas [6]

We can now write the following coupled ODE and PDE system,
where
is replaced by a continuous variable
. Differences are
replaced by derivatives and sums by integrals.

Figure 1. PolyQ41 Fibrils size distribution before (blue plain
line) and after (dashed green) 10 min of sonication. The absence
of any change in the distribution shows that neither fragmentation nor
coalescence occurred.
doi:10.1371/journal.pone.0043273.g001

[10,14] and references therein), or when applying simplifying
assumptions in the biological literature [4,6–8]. It is an efficient
tool to study protein aggregation when the average size of protein
is of a reasonable order. However, for long polymer reactions,
this modeling technique may be time-consuming and therefore
inefficient to understand the underlying complexity. One can
notice the resemblance between this infinite ODE model and a
coupled PDE [15].

From ODEs to PDE: a New Size-structured Model

We propose here a new size-structured model composed of two
ODEs and one PDE in the case of a large average size
of polymers - i.e.,
. The main idea is to replace the discrete size
of a polymer by a continuous variable
, in which we
have defined the small parameter
: 
In the same way,
the densities
are replaced by a continuous-in-size function
(see supplementary data S1 for more details). This model
can be derived from the infinite set of ODEs if the two following
assumptions hold.

First, for most polymer sizes
, there is only a slight difference
between what happens for
mers and for
-mers. In other
terms, even if quantities and reaction rates vary, it occurs in a
“continuous” manner, implying only slight changes from one size
i
to its neighbor sizes
 and
except for a small number of values. For instance, for
garation coefficients
, it is formalized as: There exists a constant, denoted below
Cst>0,
such that

For all
, 

This assertion allows a continuous viewpoint on the equations
for
. It also means that disruptions in the concentrations or in
the coefficients can only appear at some specific points, that will have
to be identified, and that are meaningful biologically. Though, this
assertion appears to be natural since the conformational changes
in polymers only occur at specific sizes [16]. Moreover, having a
look at experimental size distributions (Figure 1) confirms how
natural it is to view the size of polymers as a continuous quantity.

The second and quite standard assumption is that at the
beginning of the reaction, when polymer concentrations remain
small compared to monomers, polymerization is the main process,
whereas fragmentation and coalescence are secondary processes
[4,6]. This assumption can be replaced if necessary by a similar
one, such as the existence of a dominant polymerization by
mer addition, with
a relatively small oligomer. In such a case, the
polymerization terms
would be replaced by
in the
equations, and a similar treatment can apply.

We refer to supplementary data S1 for a rigorous mathematical
formulation of these two assertions. They are obtained when the
system of equations is rescaled, and this allows us to estimate the
relative contribution of each process to the overall dynamics.

Let us turn to the nucleus
. In this equation, the two assertions
make it possible to ignore the influence of fragmentation and
coalescence. Then as we are in the case where
, the
time-dependency of the equation for
is much faster than the one for
: it can be written

Hence, it is valid to suppose that it reaches its equilibrium
instantaneously, and we can replace Equation (6) by

We thus obtain the following equality, which generalizes well-
established formulas [6]

We can now write the following coupled ODE and PDE system,
where
is replaced by a continuous variable
. Differences are
replaced by derivatives and sums by integrals.
\[ k_{\text{on}}(x_0)\xi(t,x_0) = k_{\text{on}}(x_0) \frac{k_N^0 c_i^0}{k_{\text{off}} + k_{\text{on}}(x_0)c_i^0}. \] (12)

Complete rigorous mathematical derivation can be found in supplementary data S1, and also shows that generally the third term in the right-hand side of Equation (10) (the ratio \( \frac{\theta_0 k_N^0 (c_i^0)^{y+1} + \theta_b}{k_{\text{off}} + k_{\text{on}}(x_0)c_i^0} \)) is negligible. Even mathematical approximation theorems can be written to validate the model, as is done for instance in [10,11,17].

The advantages are twofold. First, it allows us to investigate numerically, using standard and well-known numerical schemes (see [18]), how a change in the coefficients can influence the overall reaction, and, more specifically, the size distribution. Also, inverse problem techniques could allow size-dependent parameters to be estimated (see for instance [19,20]). Secondly, it is easier to handle mathematically. Theoretical analysis can help us understand the intrinsic mechanisms and formulate new paradigms [21,22].

**Application to PolyQ Polymerization**

Aggregation of polyglutamine (PolyQ)-containing proteins is responsible for several neurodegenerative disorders including Huntington’s disease. We have carried out biophysical analyses to investigate the aggregation kinetics of PolyQ41, which are peptides containing a repetition of 41 glutamine residues per monomer. Such a length of PolyQ repetition per molecule is sufficient to induce aggregation in vitro and in transfected cells [23].

Due to its simplicity, PolyQ provides an excellent model system to test our mathematical model. According to the experimental observations (Figure 1), fragmentation can be ignored. Indeed, in Figure 1, the size distribution of PolyQ41 fibrils did not change after 10 min of ultrasound treatments, showing that polymer-to-polymer reactions do not occur.

In order to determine whether coalescence occurs, we monitored simultaneously two types of measurements, polymer size and total polymerized mass. Polymer size was estimated by a static light scattering (SLS) signal. SLS is governed by the weighted average mass of oligomers and therefore highly depends on oligomer size. It can be viewed as a measurement of

\[ I_2(t) = \sum_{i \geq b} \int c_i = \int x^2 \xi(t,x)dx. \]

Total polymerized mass was followed by thioflavine T (ThT) fluorescence. Such fluorescence arises from interactions between ThT and the peculiar structure of amyloids, relatively independently of amyloid size (above a certain size threshold). ThT can be mathematically expressed by

\[ M(t) = \sum_{i \geq b} \int c_i = \int \xi(t,x)dx. \]

If there were coalescence, the weighted average polymer size would continue to grow even when the total polymerized mass \( M(t) \) reached a plateau, so the second moment \( I_2(t) \) would continue to grow after the plateau has been reached by \( M(t) \). Here, however, both curves reach the plateau roughly simultaneously (see supplementary data S2). Therefore we conclude that coalescence is negligible. As described in Materials and Methods, the spontaneous polymerization of PolyQ41 is prevented by a glutathione s-transferase (GST) tag attached to PolyQ41 peptide. Such experimental system has the advantage of providing a system where only monomeric species are present at time 0, i.e., no seeding was required for polymerization: \( c_1(t=0)=\xi_{\text{tot}}, c_i^0(t=0)=c_i(t=0)=0 \). As the GST-polyQ41 does not constitute the pro-aggregative conformer, the PolyQ41 aggregation needs to be ignored by an irreversible enzymatic cleavage (here by thrombin hydrolysis), releasing the GST region apart from PolyQ41. This enzymatic cleavage can be assimilated to an activation process along which the poly Q41 monomer turns into a structurally activated form prone to aggregation. This led us to establish a minimal activation scheme in which the GST-polyQ41, denoted by \( c_1 \), is converted into an active form denoted \( c_i^1 \) with a constant rate \( k_i^+ \). The nucleus size \( \theta_0 \) of unknown value, can be equal to 1, 2, 3 or even more. With these assumptions, Model (4)–(7) becomes

\[ \frac{da}{dt} = -k_1 c_1 + k_i c_i^1, \]

\[ \frac{dc_1}{dt} = k_i^+ c_1 c_i^1 - k_i^- c_i^1 - \theta_0 k_N^0 (c_i^0)^{y+1} + \theta_b k_N^0 c_{\text{on}} - c_1 \sum_{i > 0} k_i^+ c_i, \]

\[ \frac{dc_0}{dt} = k_N^0 (c_i^0)^{y+1} - k_{\text{off}} c_0 - \theta_0 k_N c_0 c_i^0, \]

\[ \frac{dc_i}{dt} = c_i^1 (k_i^0 c_i - k_i^- c_i), \]

and we use the continuous version of this model, given by (9)–(12), which becomes

\[ \frac{dc_i}{dt} = -k_i c_i + k_i^+ c_i^1, \]

\[ \frac{dc_1^+}{dt} = k_i^+ c_1 c_i^1 - \theta_0 k_N^0 \frac{k_N c_i^0 (c_i^0)^{y+1}}{k_{\text{off}} + k_N c_i^0} - c_1 \int_{c_0}^{\infty} \left. k_{\text{on}}(x)c(t,x)dx \right|_{c_0}, \]

\[ \frac{\partial \bar{c}}{\partial t} = -c_i \frac{\partial}{\partial x} (k_{\text{on}}(x)c(t,x)), \]

\[ k_{\text{on}}(x_0)c(t,x_0) = k_{\text{on}}(x_0) \frac{k_N^0 (c_i^0)^{y+1}}{k_{\text{off}} + k_{\text{on}}(x_0)c_i^0}. \]

As an initial approach, we tested piecewise linear polymerization rates. They are linear from \( k_{\text{on}}^\text{min} \) to \( k_{\text{on}}^\text{max} \) on \((x_2,x_1)\), constant equal to \( k_{\text{on}}^\text{max} \) on \((x_1,x_2)\) and linearly decreasing to zero on \((x_2,x_M)\) with \( k_{\text{on}}^\text{min} \) and \( k_{\text{on}}^\text{max} \) parameters to be calibrated. We arbitrarily set \( k_{\text{on}}^\text{min} \) and \( x_1 \), which led to 7 free parameters. We have also tested two different kinds of kinetics when \( \theta_0 = 1 \): first, the special case where there is no nucleus, i.e., the polymerization process starts directly from \( c_1 \), which means \( k_{\text{on}}^\text{min} = k_{\text{on}} \) and \( k_{\text{off}} \) is negligible. This reaction scheme was unable to fit properly even a single experimental curve so we abandoned it. Second, the case when the previous model is unchanged but where \( \theta_0 = 1 \) : this
means that the nucleus \( c_{i_0} = c_{i} \) is a monomeric species differing only from \( c_1 \) in its conformation. The elongating species remains the intermediate \( c_i \). In the following, \( i_0 = 1 \) refers to this second case.

The parameters of this model were then estimated by fitting experimental data on PolyQ41 protein polymerization. We performed this in two successive ways. The first consists in fitting separately each experimental curve, corresponding to a given experiment, at a given concentration. The result is that whatever \( i_0 \) is, the fit is excellent for any curve, with a measurement error from 0.5 to 2% in \( L^2 \) adimensional norms (see supplementary data S2). It gives almost undistinguishable curves. However, the variability among the optimal coefficients was large, which led us to the second step. This consisted in fitting simultaneously all the curves of experiments carried out in identical experimental conditions, but for different concentrations. The global adimensional error (in \( L^2 \) norm) diminished with \( i_0 \), and reached its lowest level for \( i_0 = 1 \), as shown in Figure 2. For larger values of the nucleus, the error is moreover too large for the model to be acceptable (results shown in supplementary data S2). It gives solid ground to the assumption, already suggested in the literature [24], that the nucleus is of size 1, but with a specific and unconventional nucleation-elongation reaction scheme, where the elongating species \( c_1 \) and the nucleus \( c_{i_0} = c_{i} \) are distinct conformers.

Another result of our simulations is that \( k_I \) is negligible, thus we can suppose that \( c_1 = c_{0} e^{-k_I t} \). In the same way, we can compare \( c_i \) to the solution of the following differential equation

\[
\frac{dc_{test}}{dt} = k_I c_0 e^{-k_I t} - i_0 k_{on} c_0 e^{-k_I t}, \quad c_{test}(0) = 0,
\]

i.e., neglect the contribution of polymers in the equation for \( c_1 \); it fits perfectly for the total duration of the lag phase.

**Application to the Knowles et al. Model [7]**

As seen for the application to PolyQ, the fully general model (9)–(12) is not yet directly applicable, precisely because of its general character. It can be thought of as the departure point for numerical, biological and mathematical analysis; and it is indeed a powerful way to tackle polymerization issues. To illustrate our approach, we have applied our model to experimental data of amyloid protein aggregation from other authors and we have compared or transposed our model to the recently published models that were accompanying the data [7,8].

In [7], Knowles and coauthors set up a model for polymerization of breakable filament assembly. This model is an analytical approximation that they have applied to (potential) experimental data and compared to exact equations representing the experimental data. For their approximation model, Knowles and coauthors made the following assumptions.

- Polymerization at a constant rate independent of the size of the polymers,
- no degradation of polymers neither monomers,
- the size of the nucleus is \( i_0 = 2 \),
- fragmentation rate depends linearly on the size of the polymer: \( k_{off}^i = k_{off} \) constant.

![Figure 2. Simulation vs Experiments for Experimental Set 1, for an initial PolyQGST concentration of 100 \( \mu M \). The parameters were first estimated for an experimental set of initial concentration 285 \( \mu M \), then we compared the experimental measures (dotted lines) for an initial concentration 100 \( \mu M \) with the simulations (in solid lines) for \( i_0 = 1, 2, 3, 4 \). We see that the smaller \( i_0 \) is, the closer the simulation to experimental curves.](doi:10.1371/journal.pone.0043273.g002)
In the presence of the assumptions of negligible coalescence and nucleation disaggregation, the kinetics of the system can be described by the following integral equations:

\[ \frac{dP}{dt} = k_{off}(M - (2i_0 - 1)P) + k_{on}^N(C_0 - M)^0, \]

\[ \frac{dM}{dt} = (k_{on}(C_0 - M) - i_0(k_0 - 1)k_{off})P + i_0k_{on}^N(C_0 - M)^0 \]

where \( M = \sum_{i \geq 0} ic_i \) represents the total polymerized mass, and \( P = \sum_{i \geq 0} c_i \) represents the total number of polymers. These equations approximate the system by an analytical formula, justified by a fixed point method and shown numerically to give a good approximation. To apply our method, we first look at the average size \( i_M(t) \) of polymers, which is given by \( i_M = \frac{M(0)}{P(0)} \). It is shown in Figure 3 for the parameter values \( k_{off} = 10^5 M^{-1} s^{-1} \), \( k_{off} = 2.10 \times 10^5 s^{-1} \), \( C_0 = 5.10 \times 6 M \), \( k_{on}^N = 2.10 \times 10^5 M^{-1} s^{-1} \), \( i_0 = 2 \), \( M(0) = P(0) = 0 \). All these values, taken from [7], allow us to observe the evolution of the experimental curve (see Table S2 in Supplementary data S3). To overcome these limitations, we developed the following approach. Based on a large data set of experimental growth curves, transitional general parameters of the time-curve, namely the length of the lag phase \( T_{lag} \) and the slope \( k \) of the reaction curve at the inflexion point were extracted. Several theoretical models are simulated using the ODE formulation and the theoretical transitional parameters of the data. This powerful approach is based on the simulation of a full ODE system (with one equation per size of aggregates) for each model investigated and no simplifications were made to reduce the dimension of this system. As a consequence, the method is time-consuming, which limits the number of mechanisms studied and the maximal polymer size (2400 in [8]). In addition, estimation of the best fitting model is based only on general parameters of the curve, which do not seem much sensitive to the distribution of the fragmentary process (see supplementary data S3). To overcome these limitations, we propose transposing their approach using PDE models, allowing for faster simulations, no limitation in the size of aggregates, and development of inverse problem techniques ([26,27]) to estimate parameters using the overall time evolution process.

\[ k_{on} c_i x = k_{on}^N i_0^{-1} \cdot \]

If we take as in [7] \( k_{off} \) and \( k_{on} \) constant, we recover System (22) by summation, but with the terms \( (2i_0 - 1)P(t) \) and \( i_0(k_0 - 1)k_{off} \) neglected. Numerical simulations are shown in Figure 3, and we see that this simplification allows a perfect fit with the complete model, fast simulations, and a better understanding of which reaction dominates at any moment (since we have access to size distributions, see Figure 4).

**Comments on Size Distributions.** For the size parameters taken from [7], fig. 1, we are able to observe the evolution of polymer size distributions over time: see Figure 4. At the beginning of the reaction (in this particular case, for a time between 0 and 5 hours), the average size increases very fast. Then it reaches an equilibrium, and between 6 to 15 hours it reaches an exponential regime during which the whole size distribution, not only the average size, is quite steady. An explanation for this could be taken from [25] for instance. After this period, the average size decreases - and ultimately, the model shows that \( M/P \rightarrow i_0 + 1 \) but this would be accomplished only after a very long period of time. A good test for the model proposed by [7] would be to check whether size distribution of polymers resembles such a one-peak distribution. If not, the assumptions would have to be relaxed, e.g. by taking variable coefficients [25].

**PDE Model Applied to the Xue et al. Model [6]**

Xue and colleagues present a new strategy to analyse the self-assembly of misfolded proteins into amyloid fibrils [8]. They analysed fibril length distribution of \( f \)-microglobulin, a protein involved in dialysis-related amyloidosis. Xue and colleagues have developed the following approach. Based on a large data set of experimental growth curves, transitional general parameters of the time-curve, namely the length of the lag phase \( T_{lag} \) and the slope \( k \) of the reaction curve at the inflexion point were extracted. Several theoretical models are simulated using the ODE formulation and the theoretical transitional parameters of the data. This powerful approach is based on the simulation of a full ODE system (with one equation per size of aggregates) for each model investigated and no simplifications were made to reduce the dimension of this system. As a consequence, the method is time-consuming, which limits the number of mechanisms studied and the maximal polymer size (2400 in [8]). In addition, estimation of the best fitting model is based only on general parameters of the curve, which do not seem much sensitive to the distribution of the fragmentary process (see supplementary data S3). To overcome these limitations, we propose transposing their approach using PDE models, allowing for faster simulations, no limitation in the size of aggregates, and development of inverse problem techniques ([26,27]) to estimate parameters using the overall time evolution process.

Xue et al investigated \( f \)-microglobulin growth, using models including different processes: a pre-polymerization step (characterized by either no pre-polymerization, or monomer-dimer equilibrium and dimer addition mechanism, or conformation exchange), an elongation of the aggregates following a one-step function, a linear function or a power function, and a possible

\[ \frac{dc_i}{dt} = -i_0 k_{on}^N c_i^0 - c_i \int_0^\infty k_{on}(x)c(t,x)dx, \]

\[ \frac{\partial c}{\partial t} = -c \frac{\partial}{\partial x}(k_{on}(x)c(t,x)) \]

\[ + 2 \int_x^\infty k_{off}(x,y)c(y)dy - k_{off}(x)c(t,x) - k_{on}(x)c(t,x), \]
secondary process such as fragmentation. Their best-fit model is given by:

No conformational exchange, no coalescence and no degradation of polymers or monomers,

the size of the nucleus is \( i_0 = 2 \) and nucleus dissociation occurs only through depolymerization,

polymerization and depolymerization follow a one-step function with the step at \( i = 6 \),

fragmentation into two smaller polymers occurs.

Thus, using the previously introduced notations, the original ODE system can be written

\[
\frac{dc_1}{dt} = -i_0 k_{on}^N c_0^N + i_0 k_{off}^N c_0 - c_1 \sum_{j \geq 0} k_{on}^{j} c_j, \tag{26}
\]

\[
\frac{dc_{j0}}{dt} = k_{on}^N c_0^N - k_{off}^N c_{j0} - k_{on}^N c_1 c_{j0} + 2 \sum_{j \geq 0+2} k_{off}^{j0} c_j \tag{27}
\]
\[
\frac{dc_i}{dt} = c_i(k_{on}^{-1}c_i - k_{off}c_i) - (k_{dep}c_i - k_{dep}c_i + 1) + 2 \sum_{j \geq 1} k_{ij}^d c_j - K_{off}c_i. 
\]  

(28)

For the particular choice of fragmentation made in [8], however, fragmentation in polymers of size 1 is close to 0. This ODE system is then formally equivalent to the following PDE system:

\[
\frac{dc_i}{dt} = - \frac{k_{on}^N k_{off}(x_0)k_{i}^{q+1}}{k_{off} + k_{on}(x_0)c_1} - c_i \int_{x_0}^\infty k_{on}(x)c(t,x)dx, 
\]  

(29)

\[
\frac{\partial c(t,x)}{\partial t} = -c_i \frac{\partial}{\partial x}(k_{on}(x)c(t,x)) + \frac{\partial}{\partial x}(k_{dep}(x)c(t,x))' \int_{x}^\infty \frac{\partial c(t,y)}{\partial y} - K_{off}(x)c(t,x), 
\]  

(30)

\[
c(t,x_0) = \frac{k_{on}^{N}c_{i0}}{k_{off}^{i} + k_{on}(x)c_{i1}}. 
\]  

(31)

Due to the shape of the polymerization process, which has a step at \(i = 6\) (meaning that \(K_{on}^{i} = K_i\) for \(i \leq 5\), \(k_{on}^{i} = K_i\) for \(i \geq 6\)), if the step is high, that is if \(K_2 > K_1\), it is however preferable to keep all the ODEs occurring for \(i \leq 6\) and to set up the PDE (30) only for \(i \geq 6\). We then adapt the boundary condition [31] as shown in supplementary data S3. This can also be approximated by the Bishop and Ferrone model [6] by adjusting a nucleus critical size to \(b_0 = 6\). Similar work can be done for the different processes studied in [8]. Our study allowed us to enhance their approach by quick investigation of different fragmentation kernels, showing that the shape of the fragmentation does not influence the polymerization dynamics (see supplementary data S3).

Discussion

We proposed a new model (9)-(12) to serve as a global framework to investigate the mechanisms of nucleation-elongation processes in amyloid fibrils’ assemblies. We applied it to PolyQ41 aggregation, demonstrating experimentally that coalescence and fragmentation were negligible, and predicting by our simulations that the monomer activation was irreversible. Moreover, it highlighted the early step of PolyQ41 nucleus formation and assemblies. With regard to the bibliography, the concept of nucleus in protein aggregation remains obscure. Here the analysis of PolyQ polymerization suggested a kinetic scheme in which \(c_i^2\) is at an equilibrium with \(c_i\). These two species are monomeric and only differ in their conformation. According to the conventional model of nucleation-elongation process [4], the nucleus is thermodynamically stabilized by the addition of at least one monomer. Here we proposed an unconventional mechanism of nucleation in which the \(c_1\) formation constitutes the limiting step in the polymerization process which is stabilized by an interaction with \(c_i^2\). Therefore, the formation of the [\(\bar{c}_1 - c_i^2\)] complex constitutes the first proaggregative species. Furthermore, during the formation of this complex, a structural information exchange should occur between \(\bar{c}_1\) and \(c_i^2\). To reach the formation of a nucleus, two changes of conformation are hence required. The first one arises from the GST-cleavage of \(c_1\) to a conformer \(c_i^2\) released as a random coil structure, that is not proaggregative. The second change of conformation is an internal change of the random coil \(c_i^2\) into a proaggregative species \(\bar{c}_1\) that is still monomeric.

Our approach also proved highly efficient when applied to previously designed models [7,8], where it can be adapted and used to pursue the research further. We believe it could be applied to many other cases, providing both a unified framework and an efficient way to carry out fast simulations, model discrimination [28], inverse problem methods and analysis.

Materials and Methods

Model Derivation

To derive the continuous model, we first write a rescaled version of the model, that makes use of typical orders of magnitude. Then, quantifying our assumptions, we approximate sums by integrals and differences by derivatives. Finally, from the equation for \(c_i\), we deduce the boundary condition for \(c(t,x = x_0)\) (full details in supplementary data S1).

Numerical Implementation

To avoid useless conversions, we implemented the PDE model (9)-(12) with dimensioned numbers, and checked a posteriori that the considered orders of magnitude fit the assumptions. We use an explicit upwind scheme - finer methods can be used such as WENO [18].

Parameter Estimation

The parameter estimation was performed by a least-square approach. For \(b_0 = 1,2,3,4\), we searched for the optimal set of parameters such that it minimized the quadratic distance between the data points obtained by ThT measures and the simulated curve of the mass, represented by \(\int c(t,x)dx\) in the PDE model or by \(\sum c_i\) in the ODE one. The minimization task was performed by the CMAES algorithm [29]. It was run with 50 different initial parameters sets. Then the optimal solution was used as an initial guess and the minimization algorithm was run again 50 times.

Experimental Results

GST-PolyQ production. The GST-Q41 expression vector was described by Masino et al [30]. GST-polyQ41 fusion protein was produced in E.Coli BL21DE3 and purified by affinity chromatography using Glutathione Sepharose affinity beads (Pharmacia).

Fragmentation experiments. The Fragmentation experiments were performed using an immersion sonotrod oscillating at 40 Khz. The size distributions of polyQ fibrils were monitored before and after sonication by dynamic light scattering (DLS, Wyatt).

Kinetic experiments. All polymerization experiments were performed at 33°C. Aggregation was initiated by thrombin addition (0.5 unit/μM of GST-PolyQ41) leading to the release of PolyQ41 peptide from GST. The aggregation was monitored either by Thioflavine T (ThT) (100 μM) in a 96-well plate fluorescence spectrometer or by a homemade multiwavelength static light scattering/fluorescence system (SLS).
Supporting Information

Figure S1 Parameter estimation considering each curve separately. Time evolution of PolyQ41 polymerized mass for an initial PolyQGST concentration equal to 285 μM. The experimental results are plotted in dotted line and the best-fit curve in solid line. b0 is set to 3. Best-fit parameters are k+ = 0.67 h−1, k− = 0, kN = 7.8.10 2 M−2h−1, koff = 5.10 −2 h−1, kmax = 1.2.10 9 M−1h−1, x0 = 2.10 6, x3 = 0.2 i N .

Figure S2 Parameter estimation for Experimental Set 1 when b0 is set to 3. Time evolution of the adimensioned PolyQ41 polymerized mass for an initial PolyQGST concentration equal to 100 μM (A), 285 μM (B), 420 μM (C). Dotted curves represent experimental results. The solid curve is the best-fit. The global error in L^2 adimensioned norm was equal to 40% and the optimal parameters are very close to those of Figure 1.

Figure S3 Parameter estimation for Experimental Set 1 when b0 is set to 1. Time evolution of the adimensioned PolyQ41 polymerized mass for an initial PolyQGST concentration equal to 100 μM (A), 285 μM (B), 420 μM (C). Dotted curves represent experimental results. The solid curve is the best-fit. The global error in L^2 adimensioned norm was equal to 11%. The best-fit parameters are k+ = 0.65 h−1, k− = 0, kN = 7.10 −6 M−1h−1, koff = 5.10 −2 h−1, kmax = 2.3.10 9 M−1h−1, x3 = 0.1 i N , x = max = 5.10 6.

References