Kinetics of spontaneous filament nucleation via oligomers: insights from theory and simulation

Anđela Šarić*,[1,2] Thomas C. T. Michaels*,[1,3] Alessio Zaccone,[4] Tuomas P. J. Knowles,[1] and Daan Frenkel[1]

[1]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, U.K.*

[2]*Department of Physics and Astronomy, Institute for the Physics of Living Systems, University College London, Gower Street, London, WC1E 6BT, U. K.*

[3]*Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.*

[4]*Department of Chemical Engineering, University of Cambridge, Pembroke St, Cambridge CB2 3RA, U.K.*

(Dated: September 8, 2016)

Abstract

Nucleation processes are at the heart of a large number of phenomena, from cloud formation to protein crystallization. A recently emerging area where nucleation is highly relevant is the initiation of filamentous protein self-assembly, a process that has broad implications in many research areas ranging from medicine to nanotechnology. As such, spontaneous nucleation of protein fibrils has received much attention in recent years with many theoretical and experimental studies focussing on the underlying physical principles. In this paper we make a step forward in this direction and explore the early time behaviour of filamentous protein growth in the context of nucleation theory. We first provide an overview of the thermodynamics and kinetics of spontaneous nucleation in protein filaments in the presence of one relevant degree of freedom, namely the cluster size. In this case, we review how key kinetic observables, such as the reaction order of spontaneous nucleation, are directly related to the physical size of the critical nucleus. We then focus on the increasingly prominent case of filament nucleation that includes a conformational conversion of the nucleating building-block as an additional slow step in the nucleation process. Using computer simulations, we study the concentration dependence of the nucleation rate and we find that, under these circumstances, the reaction order of spontaneous nucleation with respect to the free monomer does no longer relate to the overall physical size of the nucleating aggregates but rather corresponds to the sub-aggregate size that directly participates in the conformational conversion. Our results thus provide a novel interpretation of the kinetic descriptors of protein filament formation, including the reaction order of the nucleation step or the scaling exponent of lag times, and put into perspective current theoretical descriptions of protein aggregation.

I. INTRODUCTION

Nucleation is the initial step in the formation of a new ordered structure through self-organization. It is characterized by the presence of a free energy barrier to form the smallest growth-competent unit of the new structure. Many phenomena in nature, science and engineering are nucleated processes. Familiar examples from everyday life include cloud formation, ice crystallization, the boiling of water, or the formation of bubbles in a champagne glass.

Besides these everyday systems, nucleation processes are also of fundamental importance to current research. A particularly intriguing example in this context is the formation of protein filaments, a fundamental form of biological self-assembly with important implications in areas ranging from medicine to materials science. Biofilaments of actin and tubulin for instance underlie key events in cellular life, such as providing the rigidity of the cellular cytoskeleton or participating in cell motility and cell division.[1–7] On the other side, aberrant filamentous protein aggregation is associated with over 50 increasingly prevalent human disorders, such as Alzheimer's, Parkinson's diseases and type II diabetes.[8–14] These pathologies are intimately associated with the formation and deposition in the brain or other organs of fibrillar protein aggregates, commonly known as amyloids, which are the result of the aggregation of normally soluble and functional proteins into elongated fibrillar structures characterized by their β-sheet rich structure. Amyloids, however, are not only associated with disease, as it was initially believed, but have been increasingly found to serve also many functional roles within living organisms[15] and this natural use of the amyloid state of proteins and peptides for functional purposes has inspired many applications of these structures as materials for nanotechnology.[16–20]

The formation of protein filaments has been established to be a nucleated polymerization process where a slow spontaneous (or primary) fibril nucleation step is followed by rapid growth through filament elongation[6,23,24,26,35,38–47,49–52,59,61,84] and, in certain cases, self-replication through secondary pathways.[21–25,93] When the random formation of the smallest growth-competent aggregates (nuclei) occurs directly from solution, without the participation of surfaces or nucleation seeds[21–25,93], we denote it as "spontaneous". A particularly useful approach to understand the way in which soluble proteins are converted into their fibrillar counterparts through spontaneous nucleation is represented by kinetic models of filamentous

growth[6,21,26–29]. These kinetic models allow the underlying molecular-level mechanisms of fibril formation to be connected with *in vitro* experimental measurements of the aggregate mass concentration e.g. by fluorescence microscopy or related techniques.[86] In these models, the spontaneous fibril nucleation step is commonly described as an n_c-th order reaction with respect to the free monomer concentration c with rate

$$r = k_n c^{n_c}, \tag{1}$$

where k_n is the rate constant for spontaneous nucleation and n_c is an effective reaction order of spontaneous nucleation with respect to c. Because nucleation is slow compared to growth, the value of n_c can be obtained experimentally from the concentration dependence of the half-polymerization time $t_{1/2}$ (defined as the time at which half of the monomers mass is sequestered in aggregates), as the slope of this relationship in a double logarithmic plot gives the so called scaling exponent γ defined by:

$$t_{1/2} = A \ c^{\gamma}, \quad \gamma = -n_c/2. \tag{2}$$

A key question, in the field, is how to relate these kinetic descriptors, including the reaction order n_c and the scaling exponent γ, with the microscopic characteristics of the underlying nucleation step, such as the physical size of the nucleating aggregates, thus providing important insights into the nature of the nucleation process. This problem has already received significant attention in the protein aggregation literature.[?] (cite all papers that speak about nucleation...if we order them well we should be able to have them all next to each other...)

In this paper, we make a step forward in this direction. We start by providing a brief overview of the simplest case of fibril nucleation by direct polymerization of monomers, a situation that has been considered extensively in the literature (cite all the papers of the first part), incorporating the published theories and quantitative experiments. We then consider the increasingly evident process of protein nucleation which includes an additional conformational change in the nucleating protein aggregate, giving rise to multi-step nucleation processes via small oligomers. We study the kinetics of such a process using coarse-grained computer simulations, and provide a novel interpretation of the related kinetic parameters that are commonly measured in experiments. In particular, we investigate the meaning of the reaction order n_c when proteins undergo a conformational change during nucleation, and find that n_c directly corresponds to the portion of the oligomer size that directly participates in the conversion step.

II. KINETICS OF SPONTANEOUS FIBRIL NUCLEATION WITH ONE DEGREE OF FREEDOM

We start our discussion by reviewing the simplest model of fibril nucleation, in which aggregates of different sizes but same structure are formed by direct polymerization of protein monomers. We demonstrate that this model gives rise to spontaneous nucleation if the cluster free energy function has a maximum as a function of cluster size. Under these circumstances, it is found that, independently of the specific form of the cluster free energy function, the reaction order of spontaneous nucleation n_c is always linked to the number of monomers that compose the fibril nucleus.[89] The simplicity of this model arises from the fact that the aggregate size is the only degree of freedom in the system; in Section III we relax this assumption by considering the effect of other potentially important degrees of freedom, such as the internal structure of clusters.

A. Direct polymerization models of spontaneous fibril nucleation and the nucleation theorem

To see how spontaneous nucleation emerges from a direct polymerization model, we consider the following master equation describing the time evolution of the concentration $f(t, N)$ of aggregates of N monomers under the action of elongation and dissociation processes:[53]

$$\frac{\partial f(t, N)}{\partial t} = c\, k_{\text{on}}(N - 1)f(t, N - 1) - c\, k_{\text{on}}(N)f(t, N) \tag{3}$$
$$+ k_{\text{off}}(N + 1)f(t, N + 1) - k_{\text{off}}(N)f(t, N),$$

where c is the free monomer concentration, $k_{\text{on}}(N)$ and $k_{\text{off}}(N)$ are the (size-dependent) rate constants for the addition and removal of monomers. These rate constants are linked together by the detailed balance condition

$$c\, k_{\text{on}}(N - 1)f_{\text{eq}}(N - 1) = k_{\text{off}}(N)f_{\text{eq}}(N) \tag{4}$$

where $f_{\text{eq}}(N)$ is the cluster size distribution function at equilibrium. As detailed in Refs.[56,57], Eqs. (3) can be conveniently mapped onto a one-dimensional diffusion equation in a potential landscape by assuming that the size distribution $f(t, N)$ varies sufficiently smoothly with N

so that the continuum limit approximation applies:

$$\frac{\partial f(t, N)}{\partial t} = k_{\text{on}} c \frac{\partial}{\partial N} \left[\frac{\partial f(t, N)}{\partial N} + \beta \frac{\partial \Phi(N)}{\partial N} f(t, N) \right],$$ (5)

where $\beta = 1/(kT)$ denotes the inverse temperature (k is the Boltzmann constant) and we have introduced the cluster free energy function $\Phi(N)$ defined by the relationship

$$\frac{f_{\text{eq}}(N)}{c} = e^{-\beta \Phi(N)}.$$ (6)

If the free energy function $\Phi(N)$ has a maximum at some value N^*, corresponding to the critical nucleus size, then the rate of nucleation can be obtained from Eq. (5) using the saddle point approximation:[58]

$$r \sim k_{\text{on}} c \left(\frac{\Phi''(N^*)}{2\pi \beta^{-1}} \right) e^{-\beta \Phi(N^*)}.$$ (7)

Thus, the master equation (3) of direct polymerization yields spontaneous nucleation. An important point to recognize here is that Eq. (7) is valid for arbitrary cluster free energy functions, so that depending on the specific form of $\Phi(N)$ several models of spontaneous nucleation can be formulated. Classical nucleation theory,[36,37] for instance, describes clusters of fractal dimension d as a compact object with associated volume and surface energy terms:

$$\Phi(N) = a\sigma N^{\frac{d-1}{d}} - N\Delta\mu, \quad (d > 1)$$ (8)

where $\Delta\mu = \beta^{-1} \log(c/c_s)$ is the supersaturation, c_s the saturation concentration, σ the surface tension (energy per unit surface) of the interface between the aggregate and the surrounding solvent and a is a geometrical prefactor. The balance between unfavourable entropy contribution from the motion of free monomers and the energy from the bonds between monomers creates a barrier. As N increases, more bonds are created between monomers eventually overcoming the unfavourable entropy contributions that make small cluster unstable. The nucleus is the rate limiting step and corresponds to the point at which the free energy $\Phi(N)$ peaks (Fig.1(b)), and according to Eq. (7) the rate of nucleation is given by:

$$r \sim c \, e^{-\frac{\beta \Delta\mu N^*}{d-1}}, \quad N^* = \left(\frac{(d-1)a\sigma}{d \, \Delta\mu} \right)^d.$$ (9)

It is easy to verify that from Eq. (9) the nucleus size N^* satisfies the relationship:

$$N^* = \frac{d \log(r)}{d \log(c)} - 1 = n_c - 1,$$ (10)

where the factor -1 comes from the concentration dependence of the prefactor in Eq. (7). Equation (10) is a remarkable result: it states that the nucleus size can be obtained from the slope of a double logarithmic plot of the nucleation rate r against the monomer concentration c. Moreover, Eq. (10) provides a direct relationship linking the physical size of nuclei to the experimentally accessible reaction order n_c of spontaneous nucleation. Interestingly, Eq. (10), which was derived here for the specific cluster free energy function of Eq. (8), turns out to be a far more general result known as the nucleation theorem.[89] This theorem states that Eq. (10) is valid for arbitrary cluster free energy functions of the form $\Phi(N) = F(N) - N\Delta\mu$, so that in a nucleating system where the aggregate number is the only relevant degree of freedom, the kinetic parameters n_c and γ can always be linked directly to the physical size of the nuclei. As a final note, we remind here that, in the context of classical nucleation theory, the surface energy term $F(N)$ of one dimensional clusters $(d = 1)$ is independent of N so that the cluster free energy function $\Phi(N)$ has no maximum (Fig. 1(b)). Hence, direct polymerization in 1D is a downhill process, where every aggregate is more stable that the previous ones; there is no classical nucleation in a truly 1D system.[37]

Direct polymerization models have been used widely in the protein aggregation literature to describe spontaneous fibril nucleation. Important examples include classical nucleation theory descriptions of amyloid as elongated 2D crystals.[65–67] By considering prismatic aggregates of fixed width but variable length and thickness build up by successively layered β-sheets, expressions for the nucleation rate, nucleus size and nucleation work have been obtained. In accordance with the nucleation theorem Eq. (10), these theories predict the existence of a well-defined nucleus size which can be obtained from a log-log plot of the nucleation rate against monomer concentration. Recent simulation and theoretical studies of fibril nucleation have highlighted that introducing an interaction anisotropy in these models results in a non-standard nucleation mechanism where the concept of a well-defined nucleus size breaks and nuclei with varying size are observed instead.[71–73]

B. Pre-equilibrium models of spontaneous fibril nucleation

Worth mentioning at this point are pre-equilibrium models of fibril nucleation, which effectively correspond to direct polymerization mechanisms. These models are characterized by two main assumptions. The first assumption is that clusters do not possess internal
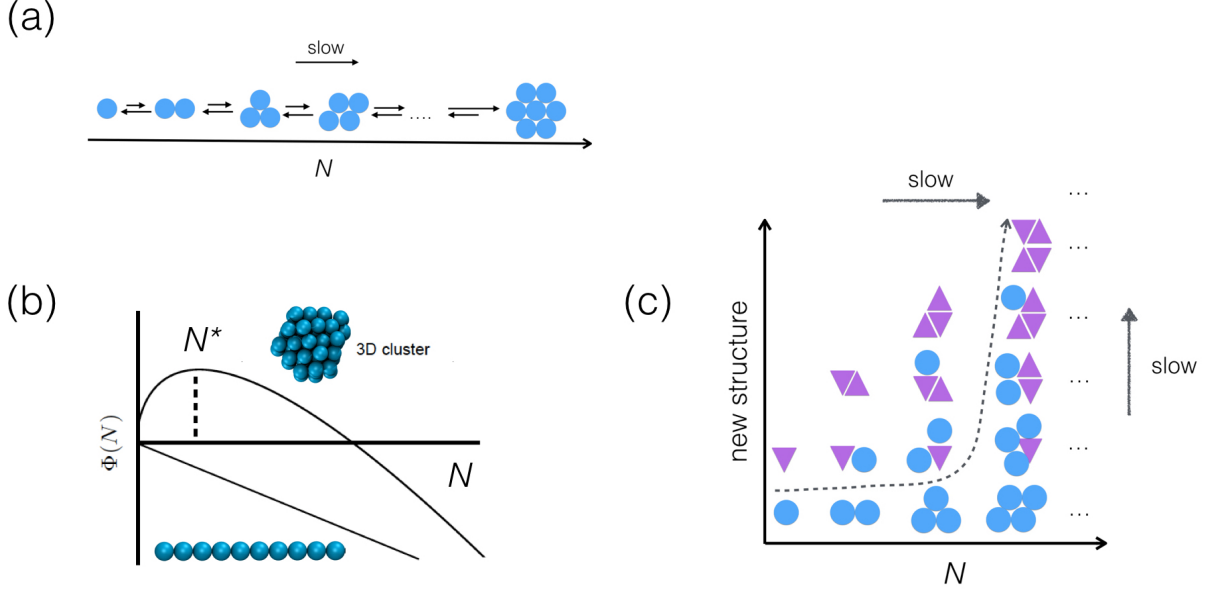
FIG. 1. (a) One step nucleation is characterised by a single slow coordinate that corresponds to the cluster size. (b) Cluster free energy of classical nucleation theory for a 3D spherical cluster and for a 1D cluster. Note that in 1D there is no free energy barrier. (c) Multistep nucleation is characterised by additional slow coordinates. In this example the additional slow coordinate is the structural change. The dashed line indicates a possible nucleation pathway.

structure, so that the nucleus can be considered as a small piece of a long aggregate and the cluster size is the only relevant degree of freedom in the system. The second assumption is that the nuclei are in equilibrium with the soluble monomer, so that their concentration can be obtained by equating the respective chemical potentials and the original kinetic problem is now transformed into an equilibrium one. In these models the rate of nucleation can be summarized as:[6,7,23,26,27,35,59–63]

$$r = A \ c^{n_c}.$$ (11)

The exact connection between the exponent n_c and the nucleus size is model dependent. For example, some authors[23,35,61–63] describe the rate of nucleation as the rate at which nuclei elongate $r = k_+ c \ [N^*]$, where k_+ is the elongation rate constant and $[N^*]$ is the concentration of nuclei. The assumption that nuclei are in equilibrium with the monomers results in $[N^*] = K_n c^{N^*}$, where K_n is the equilibrium constant for the nucleus-monomer equilibrium. Combining these expressions yields $r = k_+ K_n c^{N^*+1}$ and we find the following

relationship between the reaction order n_c of spontaneous nucleation and N^*[89]

$$N^* = \frac{d\log(r)}{d\log(c)} - 1 = n_c - 1. \tag{12}$$

Similar arguments have been employed by other authors that yield different expressions, including $N^* = n_c + 1$,[59] or $N^* = n_c$.[6,7,26,27,60] In either case, the value of the nucleus size is directly related to the slope of a plot of $\log(r)$ against the logarithm of the monomer concentration, and thus the physical size of nuclei can in principle be accessed from kinetic measurements. Several variations of these pre-equilibrium models of nucleation have also been formulated, for instance by considering different addition rates for monomer above and below N^*,[61,62,74–76] by assuming that the nucleus is formed through successive associations of small oligomers[77–80] or by including reversible association steps for aggregate sizes below N^* and considering only irreversible polymerization for $N > N^*$.[78–81]

III.   KINETICS OF SPONTANEOUS FIBRIL NUCLEATION WITH MULTIPLE DEGREES OF FREEDOM

So far, we have considered the situation when the relevant degree of freedom of the nucleating system is the physical size of aggregates. Under these circumstances, the nucleation theorem[89] provides a direct relationship between the nucleus size and the concentration dependence of the nucleation rate. This result offers a powerful strategy for accessing key information about the underlying nucleation step from experimental measurements of aggregation kinetics. We now make a step forward and consider the more complicated situation when the nucleation process is controlled by additional relevant degrees of freedom, such as the internal structure of clusters. A prominent example of such a situation is realized when the aggregating species change their shape or conformation during nucleation. In fact, when considering the assembly of soft species, such as proteins, one needs to take into account the fact that the species within the final aggregate might be in a significantly different state from their counterparts in solution. In particular, the protein conformation and shape might be substantially altered in the final aggregate. This situation can be viewed as a nucleation process that is governed by multiple degrees of freedom, in addition to the cluster size, as sketched in Fig. 1(c). An important realization of this scenario is the aggregation of amyloid fibrils, where proteins acquire a β-sheet conformation within the fibril, which is typically

very different from their native conformations in solution. A large number of structurally unrelated proteins form this type of fibrils, and amyloid fibril formation is regularly accompanied by a marked change in protein conformation. As demonstrated below, in this scenario we do not find a relationship between nucleus size N^* and the reaction order n_c, but rather find that for a given value of n_c the overall size of the nucleating oligomers can change. Interestingly, however, we find that a modified nucleation theorem connects n_c to the sub-oligomer size within which the conformational change takes place.

Amyloidogenic proteins can be characterised according to their propensity to acquire the β-sheet conformation[110–113], which controls the rate and pathways of amyloid fibril nucleation [95–101]. For proteins with low β-propensity, the conformational change from the native into the β-sheet form is slow and energetically unfavourable, and the fraction of the proteins in the β-sheet conformation in the protein aggregate acts as a second slow degree of freedom in amyloid fibril formation, in addition to the aggregation number.

A number of experimental studies have reported the existence of non-β-sheet clusters during amyloid formation[102–109], and have suggested a multi-step nucleation mechanisms, where fibril nucleation takes place via disordered prefibrillar clusters.[40,43,82–84,86] This nucleation scenario, also called a nucleated conformational conversion or a two-step nucleation, has also been in focus of several theoretical studies[68–70,85,92,95,96,99].

A.   Computer simulations of amyloid nucleation

As a quantitative understanding of aggregation processes with more than one slow degree of freedom is still presently lacking, it is beneficial to obtain interpretation of its kinetic descriptors. Coarse-grained computer simulations can be of great help in this case. Here, we use coarse-grained Monte Carlo simulations to study the kinetics and thermodynamics of nucleation of amyloid-like fibrils, for proteins with a range of β-propensities, attempting to rationalise experimentally measurable kinetic parameters in terms of the underlying microscopic steps. Amyloid fibril formation is known to involve pre-nucleation disordered oligomers, which do not posses much β-sheet content typical for fibrils, hence the oligomer aggregation number emerges as one slow degree of freedom, while the β-sheet content, typical for fibrils, emerges as the second one. Our model, although presented in the context of amyloid nucleation, is generic and can be applied to formation of any protein filaments

which involve a conformational change.

We employ a minimal computational model that reproduces fibril nucleation, as described in our previous work[92,93]. Briefly, the model accounts for the fact that amyloidogenic peptides and proteins exist in minimally two states: a state in solution (denoted "s") that can form disordered oligomers, and a higher free-energy state that can form the β-sheet enriched fibrils (denoted "β")[92,94]. The "s" state is modelled as a hard spherocylinder with an attractive patch at the tip, which accounts for non-specific interprotein interactions, and drives the formation of small disordered oligomers, as depicted in Fig. 2 (a) and (b). The strength of the attraction between two "s" proteins is characterised with a parameter $\epsilon_{ss} = 5.5\text{k}T$. The fibril forming configuration is described as a hard spherocylinder with an attractive side patch, which captures the interactions between the β-sheet prone state, such as the hydrogen-bonding and hydrophobic interactions, and leads to the fibrillar aggregates (Fig. 2 (a) and (b)). The magnitude of this attraction is considerably stronger than between the soluble state, with $\epsilon_{\beta\beta} = 30\text{k}T$. The "$s$"-"$\beta$" interaction was set to $\epsilon_{s\beta} = \epsilon_{ss} + 1\text{k}T$, as in our previous work[92]. Throughout the text k denotes Boltzmann's constant and T the temperature. We start our simulations with 600 proteins randomly distributed in a periodic cubic box, with all proteins in the "s" state. A protein is randomly chosen to be swapped between the "s" and "β" state with a probability P_{swap}. The "s" $\rightarrow$ "β" swap is thermodynamically unfavourable, and is penalised with an excess chemical potential of $\Delta\mu_{s\beta}$, to reflect the loss of the conformational entropy of the β-hairpin compared to the form in solution. This value of $\Delta\mu_{s\beta}$ quantifies the protein's β-sheet propensity, and controls the additional slow degree of freedom. The degree of oligomerisation of the proteins in the "s"-state can be controlled via the protein concentration c, as probed in the text. A protein belongs to an oligomer if it is within the attraction range from at least one other protein in the oligomer, where the range of attraction specified in our earlier work[92].

To obtain information about the kinetics of fibril nucleation, we use the mean first passage time as the proxy for nucleation rate, and calculate the rate of primary nucleation as the inverse of the average lag time for nucleation[93]. The lag time is defined as the number of Monte Carlo steps needed for the first oligomer consisting of at least two β-proteins to appear in the simulation, since the appearance of such a nucleus always leads to further fibril growth in our simulations. We however note that such an oligomer can contain any composition of the proteins in the "s" state and can be of any overall size.. The average lag
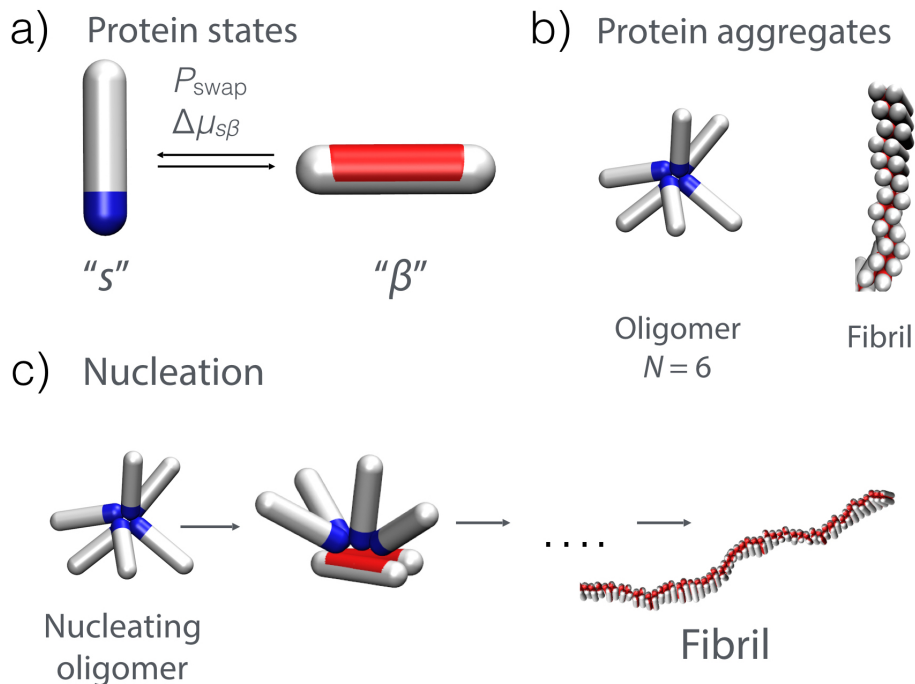
FIG. 2. The Monte Carlo model. (a) The protein can switch between two states: the soluble state, "s", that is lower in energy, interacts weakly with its own kind, and forms disordered oligomers; and the β-sheet prone state,"β", which is higher in energy, but interacts stronger with its own type than the soluble state, forming fibrils. Attractive patches are coloured in blue and red for the "s" and "β" state respectively. (b) Aggregates of the two possible states: disordered oligomer (left panel) and fibril (right panel). (c) At low protein concentrations nucleation proceeds via oligomers. Illustration of an oligomer of the size $N = 6$, where the nucleation takes place when two proteins simultaneously convert into the β-prone state, which triggers the nucleus growth into a long fibril.

time is then calculated from at least 6 repetitions of the same system with different random seeds, and is expressed in the units of 10^8 Monte Carlo (MC) steps.

B.   Kinetics of amyloid nucleation computed in simulations

At low protein concentrations, which is the regime we focus on in this paper, nucleation proceeds via oligomers, as depicted in Fig. 2(c). Starting from an equilibrated population of soluble proteins and their oligomers, we have calculated the rate of fibril nucleation across a large range of protein concentrations, as shown in Fig. 3(a), for the case of a protein

12

with a low propensity for the β-state ($\Delta\mu_{s\beta} = 20\mathrm{k}T$) and the protein with a relatively high β-propensity, $\Delta\mu_{s\beta} = 10\mathrm{k}T$. As expected, the rate of nucleation increases with the protein concentration, and with the increase in the protein β-propensity. The scaling exponent, n_c, which relates the reaction order of the nucleation step with the monomer concentration is given by the slope of the plot in Fig. 3(a), and is found to decrease with increasing β-propensity.

What is the physical meaning of the scaling exponent n_c? Its obvious characteristics is that n_c sharply decreases in the vicinity of the critical micelle concentration (cmc). This is the concentration above which increasing the total protein concentration leaves the concentration of monomers in solution unaffected (Fig. 3(b)), causing the weaker dependence of nucleation the rate on the protein concentration. Since amyloidogenic proteins are typically at very low concentrations in cells, as low as nanomolar, we focus on the meaning of the exponent at low concentration, much before the cmc is reached. The measured scaling exponent at low protein concentrations (Fig. 3(a)) is $n_c \approx 4.5$ and $\gamma \approx 2$, for the protein with a low and high β-propensity respectively.

In light of the various theories of nucleation, discussed in the Section II, where the scaling of the nucleation rate is linked with the protein concentration raised to the critical nucleus size (Eq.(12)), we measure the size (aggregation number) of the nucleating oligomer directly in our simulations. The average aggregation number of the nucleating oligomer changes between $N^* \approx 2 - 12$ across the concentration range for the low-β-propensity protein (red circles in Fig. 3(c)), and between $N^* \approx 2 - 4$ for the high-beta propensity one (blue crosses in Fig. 3(d)). Clearly, the size of the nucleating oligomer increases with the increase in the protein concentration, as predicted in our previous analysis of the free energy barriers for two-step nucleation[92]. However, for the protein with low β-propensity, the size of the nucleating oligomer does not correspond to the value of the reaction order, which is more prominent for the protein with low β-propensity. Instead, we find that the reaction order corresponds to the size of the part of the oligomer that directly participates in the conversion step, as shown by the black filled circles in Fig. 3(c) for the low β-propensity protein, and the black squares in Fig. 3(d) for the high β-propensity protein. In the latter case, conformational conversion ceases to be a slow degree of freedom, and the situation form the classical nucleation is recovered, where the reaction order corresponds to the oligomer size.

It is worth adding that in order to test for the presence of possible additional time-scales

involved in our results, we repeated the rate measurements for a larger value of the "s" $\rightarrow$ "β" conversion attempt, $P_{swap} = 1$ in our MC scheme. This choice would correspond to a protein with a very fast conformational rearrangement. In this case, the reaction was overall faster, but the scaling exponent remained unchanged (data not shown).

In what follows we set to determine what controls the value of the reaction order. As previously discussed[92], the free energy barrier for nucleation with such two slow steps is a trade-off between two opposing effects. At a constant concentration, the probability of protein conversion from a soluble into the β-sheet prone state increases with the increase in the cluster size N, while the probability of formation of an oligomer, given by $c(N)$, decreases with N. Hence, there will be some intermediate cluster size optimal for nucleation at a constant concentration, and such an optimal size will increase with the increase in the protein concentration. In our simulations we can measure the conversion probability of a protein within an oligomer of the given size per MC cycle , $P_c(N)$? , as shown in Fig. 4(a). We can also separately measure the steady-state concentration of oligomers at a certain protein concentration, $c(N)$, from simulations where proteins are not allowed to convert into the β-prone state, as shown in Fig. 4(b). Clearly, the conversion probability rate increases with increasing oligomer size. The reason for this observation is two-fold: firstly, larger oligomers contain a larger number of proteins available for conversion; secondly and more importantly, larger oligomers can have more possible binding partners to energetically stabilise the unfavourable 's' $\rightarrow$ "β" conversion. After the initial increase in $P_c(N)$, saturation is observed. On the other hand, the oligomer concentration, $c(N)$, decreases with increasing oligomer size (Fig. 4(b)). The probability rate of nucleation of a cluster of size N should then be given by the product $P_c(N) \cdot c(N)$. Fig. 4(c) shows this product versus the oligomer size for five different protein concentrations. The nucleation probability rate at a certain protein concentration, given by $P_c(N) \cdot c(N)$, clearly exhibits a maximum, which corresponds to the most probable oligomer size for nucleation, N^*. It essential to notice that this oligomer size N^* does not correspond to the most probable oligomer size observed in the system, which is $N = 2$ (Fig. 4(b)). The cluster size distribution, $c(N)$, modulated by the conversion probability rate, P_c, results in the final nucleation probability, which will depend on the functional form of P_c, hence the oligomer's geometry.

The overall rate of nucleation should then depend, up to the prefactor, on the product of
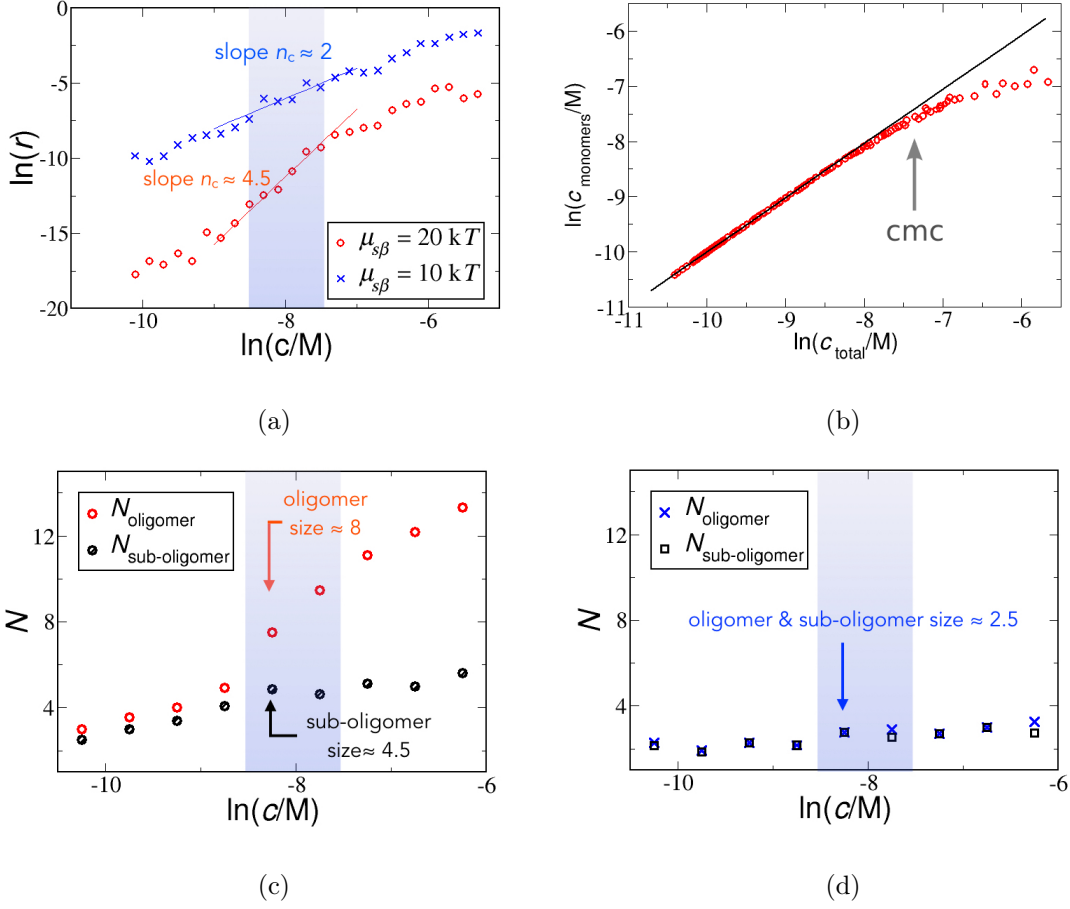
FIG. 3. Kinetics of nucleation via oligomers from Monte Carlo simulations. (a) The rate of nucleation versus protein concentration measured in simulations, for proteins with low propensity for β-sheet ($\Delta\mu_{s\beta} = 20kT$, red circle), and high β-sheet propensity ($\Delta\mu_{s\beta} = 10kT$, blue cross). (b) The concentration of free monomers in solution versus the total protein concentration. The red solid line has the slope of 1, and the dashed line indicates the critical micelle concentrations. (c) The average size of the overall nucleating oligomer (red circles) for the low β-propensity protein ($\Delta\mu_{s\beta} = 20kT$ from (a), and the average size of the sub-oligomer participating in the conversion step (black circles) . (d) The average sizes of the overall nucleating oligomer (blue cross) and the sub-oligomer participating in the conversion step (black square) for the high β-propensity protein ($\Delta\mu_{s\beta} = 10kT$ from (a).

the two contributions, summed over all possible cluster sizes:

$$r \sim \sum_{N=1}^{\infty} P_c(N) \cdot c(N). \tag{13}$$

15

(a)

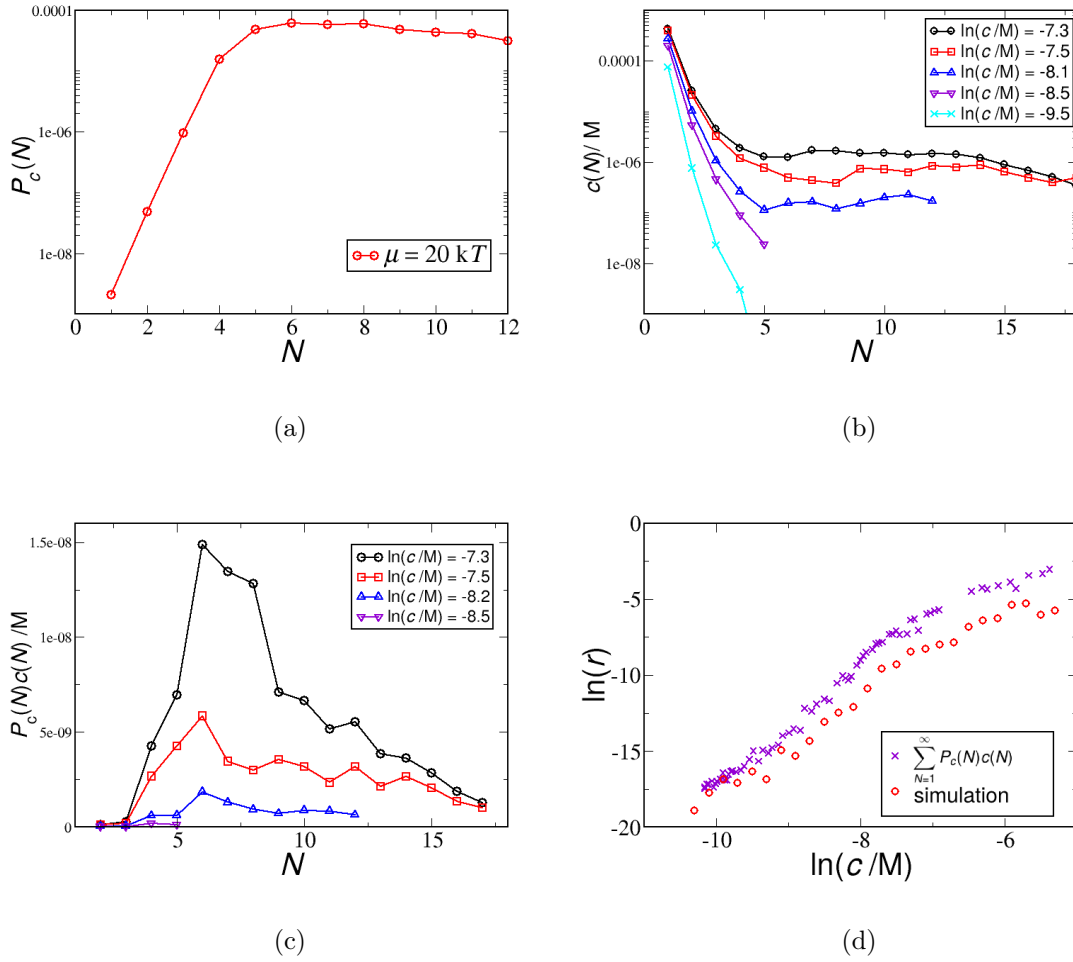(b)

(c)

(d)

FIG. 4. (a) The calculated probability of "s" $\rightarrow$ "β" conversion of a single protein within an oligomer of the size N, for the protein with a low β-propensity, $\Delta\mu_{s\beta} = 20\mathrm{k}T$, per one MC step. (b) Steady-state concentration of "s"-state oligomers of size N for five different protein concentrations. (c) Product of a conversion probability within an oligomer of a size N, from (a), with the concentration of the oligomers of size N, from (b).(d) The calculated rate: Sum of the concentrations of all oligomers multiplied by the corresponding protein conversion rates within an oligomer, across the concentration range, as in Eq.(13). The nucleation rate measured in simulations for the same set of parameters is shown for comparison (red circles), taken from Fig.3(a).

Note that the dimension of r are given in the units of concentration per MC step. Using Eq.(13), we calculated the rate of nucleation, r, for the protein with a low β-propensity $\Delta\mu_{s\beta} = 20\mathrm{k}T$ (Fig. 4(d)), and compared it to the corresponding nucleation rate measured directly in simulations. The comparison shows an excellent agreement, up to a prefactor,

16

between the calculated rate and the rate measured in simulations, giving the same scaling exponent. These results indicate that the reaction order is not only governed by the size of the nucleating cluster, as it is in the case of classical nucleation, but also by the probability of conformational conversion within such a cluster, which is an additional slow degree of freedom.

The probability for the conformational conversion is in general governed by the free energy difference between the two conformational forms, given by $\Delta\mu_{s\beta}$ in our simulations, and the interactions between the species within the oligomer, given by ϵ_{ss} and $\epsilon_{s\beta}$ in our simulations. To test the hypotheses that the scaling exponent is controlled by the functional form of $P_c(N)$, we computed the conversion probability rate $P_c(N)$ for three proteins with different β-propensities, while keeping the interaction parameters unchanged, as shown in Fig. 5(a). Trivially, the conversion probability rate increases with increasing β-propensity. However, its functional form also changes, saturating at smaller oligomer sizes for proteins with higher β-propensities, as marked by arrows in Fig. 5(a). This saturation can be viewed as the number of neighbouring proteins needed for the conversion to become sufficiently probable, which is smaller for proteins with higher β-propensity, and quantifies the cooperativity in the converting system. Fig. 5(b) shows the calculated nucleation rate, following Eq.(13), for the three proteins with different β-propensities. A decrease in the β-propensity leads to an increase in the scaling exponent. This observation is consistent with the observation of an increase in the conversion cooperativity for conversion. Hence the functional form of $P_c(N)$ has a profound effect onto the reaction order n_c , and n_c can be viewed as the number of proteins within the sub-cluster that are necessary to make the conformational conversion sufficiently probable .

CONCLUSIONS

In this paper, we have studied the phenomenon of spontaneous nucleation in the context of protein filament formation, a process which has been discussed in the literature for over 50 years and has important implications in many areas of research. We have considered the phenomenon of spontaneous fibril formation first by assuming that the cluster size is the only relevant degree of freedom in the system. In this situation, the nucleation rate is
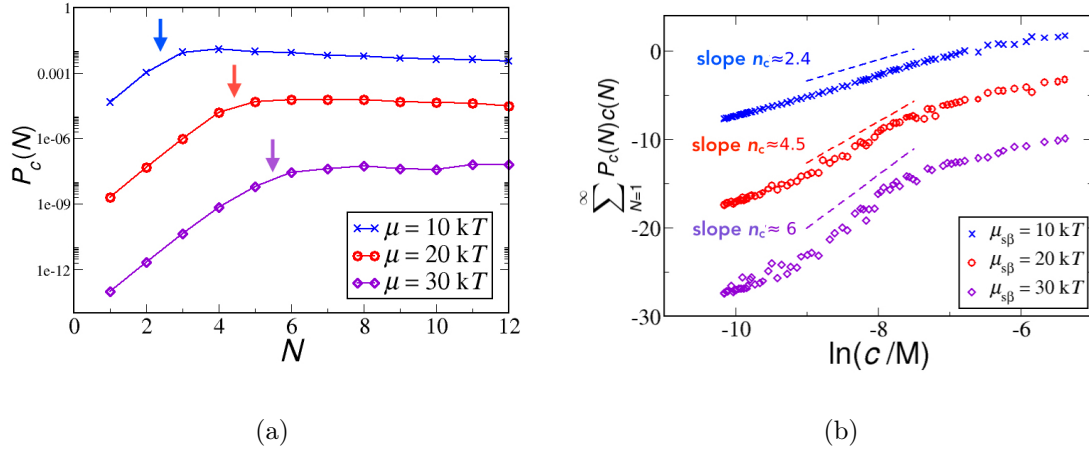
FIG. 5. (a) The calculated probability of "s" $\rightarrow$ "β" conversion of a single protein within an oligomer of the size N, for three different β-propensities, from top to bottom: $\Delta\mu_{s\beta} = 10\text{k}T$, $\Delta\mu_{s\beta} = 20\text{k}T$, and $\Delta\mu_{s\beta} = 30\text{k}T$. The arrows mark the respective saturation of the conversion probability. (b) The calculated rate, from Eq.(13), for three different β-propensities.

solely a function of the nucleus size N^*, and the kinetic descriptors commonly measured in experiments, such as reaction orders and scaling exponents, relate in a direct way with N^*.

We then introduced an additional slow coordinate in the nucleating system, namely the conformational conversion of the aggregating protein. Using coarse-grained Monte-Carlo simulations, we probed the kinetics of nucleation of protein filaments with two slow degrees of freedom. In the case of amyloid nucleation, this includes the aggregation number and the content of the fibrillar structure, the latter being characterised by the fraction of proteins in a β-sheet conformation. Our analysis showed that the reaction order in this case is not equal to the size of the aggregate, but it is modulated by the additional slow coordinate, and corresponds to the sub-cluster size that directly participates in the conversion. In particular, when conformational conversion governs the additional slow coordinate, the probability for the conformational conversion regulates the reaction order. Hence, the reaction order is related to the cooperativity for conformational conversion, and in our simulations corresponds to the sub-cluster size in which the conformation conversion becomes substantially probable. These results provide direct practical insights into the interpretation of kinetic data of fibril formation.

18

ACKNOWLEDGMENTS

---

[1] J.A. Cooper, E.L. Buhle, S.B. Walker, T.Y. Tsong, and T.D.Pollard, Biochemistry **22**, 2193 (1983).

[2] L. Tobacman and E.D. Korn, J. Biol. Chem. **258**, 3207 (1983).

[3] C. Frieden, Proc. Natl. Acad. Sci. USA **80**, 6513 (1983).

[4] C. Frieden, Annu. Rev. Biophys. Biophys. Chem. **14**, 189 (1985).

[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, Molecular Biology of the Cell, 4th edition, New York: Garland Science (2002).

[6] F. Oosawa and S. Asakura, *Thermodynamics of the polymerisation of proteins*, (Academic Press: Waltham, MA, USA, 1975).

[7] F. Oosawa and M. Kasai, J. Mol. Biol. **4**, 10 (1962).

[8] C. M. Dobson, Nature **418**, 729 (2002).

[9] C. M. Dobson, Nature **426**, 884 (2003).

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

[11] D. M. Walsh and D. J. Selkoe, *J. Neurochem.* **101**, 1172 (2007).

[12] E. Rhoades, J. Agarwal, and A. Gafni, *Biochim. Biophys. Acta* **1476**, 230 (2000).

[13] W. Dauer and S. Przedborski, *Neuron* **39**, 889 (2003).

[14] M. DiFiglia, E. Sapp, K. O. Chase, S. W. Davies, G. P. Bates, J. P. Vonsattel, and N. Aronin, *Science* **277**, 1990 (1997).

[15] Fowler, D. M. *et al.*, Trends Biochem Sci **32**, 217 (2007).

[16] M. Reches and E. Gazit, Science **300**, 625 (2003).

[17] T.P.J. Knowles, T.W. Oppenheim, A.K. Buell, D.Y. Chirgadze, and M.E. Welland, Nat. Nanotechnol. **5**, 204 (2010).

[18] T. P. J. Knowles and M. J. Buehler, Nature Nanotechnology **6**, 469 (2011).

[19] E. Gazit, *Chem. Soc. Rev.* **36**, 1263 (2007).

[20] C. Li, J. Adamcik, and R. Mezzenga, *Nat. Nanotechnol.* **7**, 421 (2012).

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

[22] A.M. Ruschak and A.D. Miranker, Proc. Natl. Acad. Sci. U.S.A. **104**, 12341 (2007).

[23] F.A. Ferrone, J. Hofrichter, H.R. Sunshine, and W.A. Eaton, Biophys. J. **32**, 361 (1980).

[24] S. Linse, C. Cabaleiro-Lago, W.F. Xue, I. Lynch, S. Lindman, E. Thulin, S. E. Radford, and K. A. Dawson, Proc. Natl. Acad. Sci. U.S.A. **104**, 8691 (2007).

[25] C. Cabaleiro-Lago, F. Quinlan-Pluck, I. Lynch, S. Lindman, A. M. Minogue, E. Thulin, D.M. Walsh, K. A. Dawson, and S. Linse, J. Am. Chem. Soc. **130**, 15437 (2008).

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

[27] S.I.A. Cohen, M. Vendruscolo, M.E. Welland, C.M. Dobson, E.M. Terentjev, and T.P.J. Knowles, J. Chem. Phys. **135**, 065105 (2011).

[28] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, J. Chem. Phys. 135, 065106 (2011).

[29] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, J. Chem. Phys. 135, 065107 (2011).

[30] S.M. Catalano, E.C. Dodson, D.A. Henze, J.C. Joyce, G.A. Krafft, and G.G. Kinney, Curr. Top. Med. Chem. **6**, 597 (2006).

[31] H.A. Lashuel, D. Hartley, B.M. Petre, T. Walz, and P.T. Lansbury, Jr, Nature **418**, 291 (2002).

[32] T. C. T. Michaels, G. A. Garcia, and T. P. J. Knowles, J. Chem. Phys. **140**, 194906 (2014).

[33] T.C.T. Michaels, P. Yde, J.C.W. Willis, M.H. Jensen, D. Otzen, C.M. Dobson, A.K. Buell, and T.P.J Knowles, J. Chem. Phys. **143**, 164901 (2015).

[34] T.C.T. Michaels, H.W. Lazell, P. Arosio, and T.P.J. Knowles, J. Chem. Phys. **143**, 054901 (2015).

[35] F. A. Ferrone, J. Hofrichter and W. A. Eaton, J. Mol. Biol. **183**, 591 (1985).

[36] F. F. Abraham, *Homogeneous nucleation theory*, Academic (New York, USA, 1974).

[37] D. Kashchiev, *Nucleation: basic theory with applications*, Oxford University Press (2000).

[38] J. Hofrichter, J. Mol. Biol. **189**, 553 (1986).

[39] J. T. Jarrett and P. T. Lansbury, Cell **73**, 1055 (1993).

[40] A. Lomakin, D. B. Teplow, D. A. Kirschner, and G. B. Benedek, Proc. Natl. Acad. Sci. U.S.A. **94**, 7942 (1997).

[41] J. D. Harper and P. T. Lansbury, Jr., Annu. Rev. Biochem. **66**, 385 (1997).

[42] O. Galkin and P. G. Vekilov, J. Mol. Biol. **336**, 43 (2004).

[43] T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. F. Arnsdorf, and S. L. Lindquist, Science **289**, 1317 (2000).

[44] H. D. Nguyen and C. K. Hall, Proc. Natl. Acad. Sci. U.S.A. **101**, 16180 (2004).

[45] H. D. Nguyen and C. K. Hall, J. Biol. Chem. **280**, 9074 (2005).

[46] H. D. Nguyen and C. K. Hall, J. Am. Chem. Soc. **128**, 1890 (2006).

[47] F. A. Ferrone, Methods Enzymol. **412**, 285 (2006).

[48] O. Galkin, R. L. Nagel, and P. G. Vekilov, J. Mol. Biol. **365**, 425 (2007).

[49] S. Auer, C. M. Dobson, and M. Vendruscolo, HFSP J. **1**, 137 (2007).

[50] S. Auer, C. M. Dobson, M. Vendruscolo, and A. Maritan, Phys. Rev. Lett. **101**, 258101 (2008).

[51] W.-F. Xue, S. W. Homans, and S. E. Radford, Proc. Natl. Acad. Sci. U.S.A. **105**, 8926 (2008).

[52] J. Zhang and M. J. Muthukumar, J. Chem. Phys. **130**, 035102 (2009).

[53] Equations (3) are commonly known as Becker-Döring equations in recognition of their work in the 1930's on nucleation;[54] the complete formulation of these equations, however, was provided only later by Penrose and Lebowitz.[55]

[54] R. Becker and W. Döring, Ann. Phys. **24**, 719 (1935).

[55] O. Penrose and J. L. Lebowitz, *Studies in Statistical Mechanics VII: Fluctuation Phenomena* ed. E. Montroll and J. L. Lebowitz (Amsterdam: North-Holland, 1976).

[56] Ya. B. Zeldovich, J. Phys. (USSR) **12**, 525 (1942).

[57] A. Zaccone, J. J. Crassous, and M. Ballauff, J. Chem. Phys. **138**, 104908 (2013).

[58] H. A. Kramers, Physica **7**, 284 (1940).

[59] J. Hofrichter, P. D. Ross, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **71**, 4864 (1974).

[60] J. M. Andrews and C. J. Roberts, J. Phys. Chem. B **111**, 7897 (2007).

[61] M. F. Bishop and F. A. Ferrone, Biophys. J. **46**, 631 (1984).

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

[63] A. Vitalis and R. V. Pappu, J. Chem. Phys. **159**, 14 (2011).

[64] D. Kashchiev, J. Chem. Phys. **129**, 164701 (2008).

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

[66] R. Cabriolu, D. Kashchiev, and S. Auer, J. Chem. Phys. **133**, 225101 (2010).

[67] R. Cabriolu and S. Auer, J. Mol. Biol. **411**, 275 (2011).

[68] A. Vitalis and R. V. Pappu, Biophys. Chem. **159**, 14 (2010).

[69] J.D. Schmit, K.Ghosh, and K. Dill, Biophys. J. **100**, 450 (2011).

[70] S. Auer, P. Ricchiuto, and D. Kashchiev, J. Mol. Biol. **422**, 723 (2012).

[71] R. Cabriolu, D. Kashchiev, and S. Auer, J. Chem. Phys. **137**, 204903 (2012).

[72] D. Kashchiev, R. Cabriolu, and S. Auer, J. Am. Chem. Soc. **135**, 1531 (2013).

[73] R. J. Bingham, L. G. Rizzi, R. Cabriolu, and S. Auer, J. Chem. Phys. **139**, 241101 (2013).

[74] E.T. Powers and D.L. Powers, Biophys. J. **91**, 122 (2006).

[75] E.T. Powers and D.L. Powers, Biophys. J. **94**, 379 (2008).

[76] R.F. Goldstein and L. Stryer, Biophys. J. **50**, 583 (1986).

[77] H. Flyvbjerg, T.E. Holy, S. Leibler, Phys. Rev. E **54**, 5538 (1996).

[78] V. Sluzky, J.A. Tamada, A.M. Klibanov, and R. Langer, Proc. Natl Acad. Sci. USA **88**, 9377 (1991).

[79] M. Kodaka, Biophys. Chem. **109**, 325 (2004).

[80] P. Arosio, S. Rima, M. Lattuada, and M. Morbidelli, J. Phys. Chem. B **116**, 7066 (2012).

[81] Y. Li and C.J. Roberts, J. Phys. Chem. B **113**, 7020 (2009).

[82] J. Lee, E.K. Culyba, E.T. Powers, and J.W. Kelly, Nat. Chem. Biol. **7**, 602 (2011).

[83] P.G. Vekilov, Nanoscale **2**, 2346 (2010).

[84] O. Galkin, W. Pan, L. Filobelo, R.E. Hirsch, R.L. Nagel, and P.G. Vekilov, Biophys. J. **93**, 902 (2007).

[85] D. Kashchiev, P.G. Vekilov, and A.B. Kolomeisky, J. Chem. Phys. **122**, 244706 (2005).

[86] E. Hellstrand et al., ACS Chem. Neurosci. 1, 13 (2010).

[87] S. Auer and D. Kashchiev, Phys. Rev. Lett. **104**, 168105 (2010).

[88] S. Auer, J. Chem. Phys. **135**, 175103 (2011).

[89] D. Kashchiev, J. Chem. Phys. **76**, 5098 (1982).

[90] D. Kashchiev, J. Chem. Phys. **125**, 014502 (2006).

[91] J. H. ter Horst and D. Kashchiev, Chem. Phys. **123**, 114507 (2005).

[92] A. Šarić, Y. C. Chebaro, T. P. J. Knowles, D. Frenkel, Proc. Natl. Acad. Sci. U.S.A. **11**, 17869 (2014).

[93] A. Šarić, A. K. Buell, G. Meisl, T. C. T. Michaels, C. M. Dobson, S. Linse, T. P. J. Knowles, D. Frenkel, Nat. Phys., in press (2016).

[94] N. S. Bieler, T. P.Knowles, D. Frenkel, R. Vácha, PLoS Comput Biol, **8**, e1002692 (2012).

[95] R. Pellarin, A. Caflisch, J. Mol. Biol., **360**, 882 (2006).

[96] R. Pellarin, E. Guarnera, A. Caflisch, J. Mol. Biol., **374**, 917 (2007).

[97] R. Pellarin, P. Schuetz, E. Guarnera, A. Caflisch, J. Am. Chem. Soc., **132**, 14960 (2010).

[98] G. Bellesia, J.-E. Shea, J. Chem. Phys., **130**, 145103 (2009).

[99] G. Bellesia, J.-E. Shea, J. Chem. Phys., **131**, 111102 (2009).

[100] C. Wu, J.-E. Shea, Curr. Opin. Struct. Biol., **21**, 209 (2011).

[101] A. Morriss-Andrews, J.-E. Shea, Ann. Rev. Phys. Chem., **66**, 643 (2015).

[102] W. Garzon-Rodriguez, M. Sepulveda-Becerra,S. Milton, C. G. Glabe CG, J. Biol. Chem., **272**, 21037 (1997).

[103] G. M. Shankar et al. (2008), Nat. Med., **14**, 837 (2008).

[104] Y. Liang Y, D. G. Lynn, K. M. Berland, J. Am. Chem. Soc., **132**, 6306 (2010).

[105] S. L. Bernstein et al. J. Am. Chem. Soc., **127**, 2075 (2005).

[106] G. Bitan et al., Proc. Natl. Acad. Sci. U.S.A. **100**, 330 (2003).

[107] C. Bleiholder ,N. F. Dupuis,T. Wyttenbach,M. T. Bowers, Nat. Chem., **3**, 172 (2011).

[108] R. Sabate , J. Estelrich, J. Phys. Chem. B, **109**, 11027 (2005).

[109] W. Yong et al., Proc. Natl. Acad. Sci. U.S.A. **99**, 150 (2002).

[110] S. Zibaee, O. S. Makin, M. Goedert, L. C. Serpell, Prot. Sci., **16**, 906 (2007).

[111] A. G. Street, S. .L Mayo. proc. Natl. Acad. Sci. U.S.A. **96**, 9074 (1999).

[112] L. Tjernberg, W. Hosia, N. Bark, J. Thyberg, J. Johansson, J. Biol. Chem., **277**, 43243 (2002).

[113] A. P. Pawar, K. F. DuBay, J. Zurdo, F. Chiti, M. Vendruscolo, C. M. Dobson, J. Mol. Biol, **350**, 379 (2005).