Published in final edited form as:
Isr J Chem. 2017 July ; 57(7-8): 738–749. doi:10.1002/ijch.201600079.

Theory of amyloid fibril nucleation from folded proteins

Lingyun Zhang*,[a],[b] and **Jeremy D. Schmit**†,[a]

[a]Department of Physics, Kansas State University, Manhattan, KS 66506, USA

[b]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China

Abstract

We present a theoretical model for the nucleation of amyloid fibrils. In our model we use helix-coil theory to describe the equilibrium between a soluble native state and an aggregation-prone unfolded state. We then extend the theory to include oligomers with β-sheet cores and calculate the free energy of these states using estimates for the energies of H-bonds, steric zipper interactions, and the conformational entropy cost of forming secondary structure. We find that states with fewer than ~10 β-strands are unstable relative to the dissociated state and three β-strands is the highest free energy state. We then use a modified version of Classical Nucleation Theory to compute the nucleation rate of fibrils from a supersaturated solution of monomers, dimers, and trimers. The nucleation rate has a non-monotonic dependence on denaturant concentration reflecting the competing effects of destabilizing the fibril and increasing the concentration of unfolded monomers. We estimate heterogeneous nucleation rates and discuss the application of our model to secondary nucleation.

Keywords

aggregation; Amyloid beta-peptides; Biophysics; Fibrous proteins; Protein models

I. INTRODUCTION

Amyloidogenic peptides have been observed *in vitro* to form a wide array of aggregate morphologies. These experiments are difficult to interpret because it is not clear which aggregation products form under physiological concentrations and which are relevant for disease progression. Insight into the former question can be obtained by mapping out an aggregation phase diagram to understand how the observed aggregation state depends on solution conditions[1,2]. The weakness of these equilibrium approaches is that often kinetic factors prevent the system from reaching equilibrium on experimental or even physiological timescales. A good example of this is the protection against aggregation provided by the natively folded state. Evolutionary pressure has limited the exposure of aggregation prone residues on protein surfaces, so aggregation requires unfolding events in multiple proteins before intermolecular association can occur[3]. Since proteins have folding stabilities on the

*lyzhang@iphy.ac.cn
†schmit@phys.ksu.edu

order of 10 kJ/mol[4], this provides a prohibitive barrier in most cases (the autocatalytic activity of prions is an important exception[5]). Other important kinetic limitations include the nucleation barrier associated with the formation of a new phase and the sequestration of proteins into off-pathway metastable aggregates like oligomers and precipitates[6].

In some cases the timescales associated with the formation of different states are sufficiently separated that pseudo-equilibrium models can predict the system behavior[6]. In other cases, multiple processes can occur on similar timescales requiring elaborate mass-action theories to disentangle the contributions of various pathways. This approach has been instrumental in elucidating the roles of fragmentation and secondary nucleation in the proliferation of fibrils following the initial primary nucleation event [7–10]. However, both of these approaches share a common limitation in that the kinetic predictions are not sensitive to variations in the solution conditions. This is an essential feature when attempting to infer physiological implications from experiments conducted under conditions that greatly accelerate aggregation.

In order to obtain the necessary sensitivity to system conditions, we require a theory that incorporates the microscopic dynamics of aggregation. A pair of useful reaction coordinates for such a theory is the number of intermolecular H-bonds and the alignment, or “registry” between neighboring molecules[11,12]. High resolution structures of mature fibrils show the constituent proteins forming in-register β-sheets resembling one-dimensional crystals[13–15]. To find this highly ordered state, aggregating molecules must sample many different registries, requiring the formation and breakage of many H-bonds[11,12]. This is a slow process, on the order of milliseconds per registry, with the result that aggregation occurs much slower than the formation of secondary structure in the folding of a single protein[12]. This search over registry states has important implications for the influence of solution conditions. At high protein concentrations the diffusion time is faster than the time required for incorrectly aligned proteins to unbind from the fibril. This means that most collisions between monomers and the fibril end cannot lead to successful growth due to the presence of incorrectly bound proteins capping the fibril end. Therefore, weakening the intermolecular bonds, either by increasing the temperature or adding denaturant, will actually increase the rate of aggregation[12]. At low concentration, the diffusion time is slow enough that incorrectly bound proteins can complete the sampling of states before the next molecule attempts to bind to the fibril. In this regime, the dominant effect of weakening the intermolecular bonds is to increase the off-rate of correctly bound molecules giving the intuitive result that fibrils grow faster under conditions where they are more stable.

In a recent paper we used the H-bond reaction coordinate to model the lag times that precede the onset of aggregation[16]. In Classical Nucleation Theory the lag time is a result of the fact that an aggregating cluster must reach a critical size before the favorable energy of binding is able to offset the translational entropy cost of being confined to the cluster[17]. This mechanism is also present in protein fibrils because the cluster must reach a minimum size of four β-strands before incoming molecules can form both the H-bonding and steric zipper interactions found in mature fibrils[18–21]. This means that the second and third β-strands to add to the cluster sacrifice translational entropy without the benefit of the full set of attractive interactions found in a mature fibril. However, in amyloid fibrils a second contribution to the nucleation barrier appears from the conformational entropy cost of extending the peptide backbones into β-sheets[2]. The magnitude of this entropic penalty is such that fibrils must reach a length of ~5 β-strands before the free energy of the fibril is lower than that of the soluble monomers[16].

Our previous theory shows that the dominant nucleation pathway is a compromise between two competing effects[16]. On one hand, pre-nucleation clusters will seek low free energy states that maximize conformational disorder. On the other hand, highly disordered clusters provide a poor binding substrate for new molecules, so highly ordered clusters are more likely to retain newly bound molecules long enough to reach a stable size. As a result, the most probable nucleation pathway goes through states where the cluster is partially ordered. This compromise allows the cluster to avoid the highest free energy states while presenting a binding surface capable of retaining new molecules for an acceptable length of time.

In this paper we extend this work to consider mechanisms that will accelerate or retard nucleation rates relative to this baseline model; the native state of the protein, the search over binding registries, and impurities or interfaces that provide heterogeneous nucleation sites. Under some conditions amyloids have been shown to assemble from pre-formed oligomers[22,23]. However, these oligomers, and the resulting fibrils, are higher in free energy than the fibrils formed via monomer pathways[6]. Moreover, the low concentrations found *in vivo* are most likely below the critical concentration required for oligomer formation and subsequent assembly[24,25]. Therefore, in the first part of the paper we model a nucleation pathway in which the nucleus grows one molecule at a time. This pathway is the most likely one at low concentration and the model provides insights into the free energy barriers that must be surmounted in any pathway. As an example of more complicated pathways, we include a discussion of heterogeneous nucleation, which shows generically how non-native contacts can alleviate the entropic barrier.

We begin with a free energy analysis of the monomer and initial stages of intermolecular β-sheet formation. These latter states are “oligomers” in the generic sense, but most likely unrelated to the metastable oligomers that have attracted interest for their toxic activity. We believe these oligomers are distinct for two reasons. First, we show that the oligomers in our model are high energy states and do not lie within a free energy basin as required for metastability. Secondly, by design our oligomers lie on the fibril formation pathway and, therefore, are stabilized by contacts that are structurally distinct from those found in toxic oligomers [26–28].

II. MODEL

A. Monomer folding equilibrium

We model the proteins as a solution of α-helix forming peptides, each containing L amino acids. In this context helix-coil theory provides a toy model for a molecule that can adopt an aggregation resistant folded state and an aggregation prone unfolded state. Importantly, the helix-coil model allows for partially unfolded states, but these states are suppressed by the cooperative two-state transitions typical of folded domains[29]. The helix-coil transition is described by two parameters, the propagation parameter, s, and the initiation parameter, σ.

We adopt the usual convention where the disordered coil state has a reference free energy of zero. The parameter s is the Boltzmann weight for a peptide unit to join an adjacent helix while σ reflects the entropic penalty required to initiate a helix. The partition function of the helix-coil model can be computed using transfer matrices[30,31]

$$Z_L = \begin{pmatrix} 1 & 1 \end{pmatrix} \cdot M^L \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \simeq \lambda_1^L \quad (1)$$

where L is the length of the protein, the transfer matrix is given by

$$M = \begin{pmatrix} 1 & \sigma s \\ 1 & s \end{pmatrix} \quad (2)$$

and λ_1 is the largest eigenvalue of the matrix

$$\lambda_1 = \frac{1}{2}[(1+s) + \sqrt{(1-s)^2 + 4s\sigma}] \quad (3)$$

The free energy of the monomer state is $-Lk_BT \ln \lambda_1$ and the fraction of amino acids in the helix state is

$$\theta = \frac{\partial \, ln \, \lambda_1}{\partial \, ln \, s} = \frac{s + \frac{s(s-1) + 2s\sigma}{\sqrt{(1-s)^2 + 4s\sigma}}}{(1+s) + \sqrt{(1-s)^2 + 4s\sigma}} \quad (4)$$

Since we are interested in the aggregation of unfolded proteins, a more useful parameter is the fraction of amino acids that are not in the folded state $\phi = 1 - \delta$.

B. Oligomerization and β-sheet formation

After proteins unfold, they become prone to aggregation via the formation of intermolecular β-sheets. In the simplest case each molecule contributes a single β strand to the final aggregate, but more complicated structures have also been observed, ranging from the hairpin motif of Aβ and IAPP[13,32,33] to the β-helix solenoid of the HET-s prion[14]. In the following, we present calculations for the assembly of single β strand and hairpin forming molecules, where the latter case presents the simplest situation where the molecules adopt a conformation allowing the formation of multiple β-strands. Quantities related to these two cases will be denoted with the subscripts “ss” (single strand) or “hp” (hairpin). In both cases we model the fibril as a bilayer consisting of two β-sheets with a steric zipper interface between them.

First, we consider the case of single strand molecules. The simplest aggregated species is a dimer which we define by the formation of intermolecular H-bonds (Fig. 1c). These H-bonds constrain a segment of each peptide into the extended β-sheet conformation. The entropic cost of this constraint exceeds the binding energy of the bonds so that the dimer has a net unfavorable free energy[16]. The protein segments not constrained by intermolecular bonds are free to adopt either folded or random coil conformations. We write the partition function for the dimer as

$$Z_{ss}^{(2)} \simeq \sum_{m_2=1}^{L} (L-m_2+1)^2 g_2^{m_2} \lambda_1^{2(L-m_2)} \quad (5)$$

$$\simeq Z_L^2 \sum_{m_2=1}^{L} (L-m_2+1)^2 \left(\frac{g_2}{\lambda_1^2}\right)^{m_2} \quad (6)$$

The summation variable m_2 denotes the number of intermolecular H-bonds. Each bond contributes a free energy $-k_B T \ln g_2$ which accounts for the favorable energy of the H-bond and the loss of conformational entropy from both chains. $(L - m_2 + 1)^2$ is the number of ways to select m_2 contiguous amino acids from each chain to form the bonds and the factors of λ_1 are the contribution from the peptide tails not participating in the H-bonds. The approximation in these formulae, and subsequent oligomer partition functions, comes from employing the long-chain limit for the free tails. The finite sums in these expressions can be evaluated analytically, however, the resulting expressions are unwieldy and contribute little to intuition.

Next, we calculate the free energy of the trimer state. There are two possible trimer states; one where all three molecules are part of a single β-sheet, and a trimer with two molecules in one β-sheet while the third molecule initiates a second sheet and forms steric zipper interactions with the first two. Although the former state has lower free energy (see below), we focus on the latter state since it provides the shortest pathway to a tetramer with two molecules in each β-sheet (rights panels in Fig. 1). This trimer to tetramer transition is the first molecular addition that provides bulk-like interactions with both H-bonds and steric zipper interactions and, therefore, we expect that it is the dominant path toward nucleation.

The partition function for the trimer is

$$Z_{ss}^{(3)} \simeq \sum_{m_3 \leq m_2} \sum_{m_2} (m_2-m_3+1)\,(L-m_2+1)^2 (L-m_3+1) g_2^{m_2} g_3^{m_3} \lambda_1^{3L-2m_2-m_3} \quad (7)$$

$$\simeq Z_L^3 \sum_{m_3 \le m_2} \sum_{m_2} (m_2 - m_3 + 1)\,(L - m_2 + 1)^2\,(L - m_3 + 1) \left(\frac{g_2}{\lambda_1^2}\right)^{m_2} \left(\frac{g_3}{\lambda_1}\right)^{m_3} \qquad (8)$$

Again, m_2 describes the number of H-bonds between the first two molecules and m_3 is the number of amino acids in β-conformation on the third molecule. The degeneracy factor in Eq. 8 has additional terms relative to Eq. 6 that describe where the third molecule inserts between the first two and the ways to choose m_3 amino acids from the third molecule. The propagation parameter g_3 accounts for the loss of conformational entropy from the third molecule as well as the favorable steric zipper interactions. Since the steric zipper requires order in both β-sheets, the summation over m_3 is limited to values smaller than the length of the first β-sheet.

For our calculations of the nucleation rate we require the population of trimers that provide a binding surface of exactly m_3 amino acids. This is given by

$$Z^{(3)}(m_3) \simeq Z_L^3 (L - m_3 + 1) \left(\frac{g_3}{\lambda_1}\right)^{m_3} \sum_{m_2 = m_3}^{L} (m_2 - m_3 + 1)\,(L - m_2 + 1)^2 \left(\frac{g_2}{\lambda_1^2}\right)^{m_2} \qquad (9)$$

The next largest aggregate is the tetramer. The fourth molecule is the first one that can form both backbone H-bonds and sidechain steric zipper interactions with the existing cluster. Since these are the same interactions present in the growth of a mature fibril, this addition must be thermodynamically favorable. Therefore, this step takes the cluster beyond the nucleation free energy barrier and will be described in the kinetic portion of the theory.

Now we consider the aggregation of hairpin forming molecules. In these systems each molecule contributes two β-strands to the aggregate. This means that half as many molecules need to be recruited to the aggregate in order to reach a stable size. It also means that bulk-like interactions begin with the addition of the second molecule. Therefore, when modeling the equilibrium distribution of pre-nucleation species, we need only consider the conversion of monomers between the folded, unfolded, and hairpin states. While the former two states are described by the helix-coil model, we still require the free energy of the hairpin. We write the partition function for this state as

$$Z_{hp}^{(1)}(m_{hp}) \simeq Z^{(1)} \sum_{m_{\text{loop}}} (L - 2m_{hp} - m_{\text{loop}} + 1) \frac{1}{m_{loop}^{3/2}} \left(\frac{g_{hp}}{\lambda_1^2}\right)^{m_{hp}} \qquad (10)$$

This expression describes a molecule that forms a steric zipper m_{hp} amino acids in length with the sequences contributing to this zipper separated by a disordered loop of m_{loop} amino acids. The formation of a closed loop incurs a conformational entropy penalty of $(3k_B T/2)\ln m_{\text{loop}}$ which results in the factor of $m_{\text{loop}}^{-3/2}$[34]. The amino acids in the zipper contribute a free energy $-m_{hp}k_BT\ln g_{hp}$, which accounts for the loss of conformational entropy and the favorable sidechain interactions. The degeneracy factor accounts for all the possible placements of the zipper along the peptide chain.

Note that the expressions in this section are sequence independent in that they assume that H-bond and steric zipper interactions can form between any pair of amino acids. We consider the opposite limit, that of strict sequence specificity, in section III B.

C. Estimation of parameters

Our model contains six parameters; s, σ, g_2, g_3, g_{hp}, and the mature fibril propagation constant g_4. In this section we constrain the parameter space using estimates of the microscopic interactions contained in these parameters. Following the work of Ghosh and Dill[29], we write the free energy of a helical amino acid as the sum of a H-bond energy and the conformational entropy loss

$$-k_BT\ln s = f_{\mathrm{HB}} + f_{\mathrm{CE}} \quad (11)$$

By fitting thermal unfolding curves, these contributions were found to be $f_{\mathrm{HB}}/k_BT = -1.91$ and $f_{\mathrm{CE}}/k_BT = \ln(6.83-1) = 1.76$, which gives a slightly favorable helix free energy of $-0.15k_BT$ and a nucleation parameter $\sigma = 0.005k_BT$[29].

The dimer propagation parameter describes the formation of one H-bond and loss of conformational entropy from two peptide units

$$-k_BT\ln g_2 = (f_{\mathrm{HB}} + 2f_{\mathrm{CE}}) \simeq 1.61k_BT \quad (12)$$

Note that this repulsive free energy does not account for the loss of translational entropy, which will be included in the grand canonical treatment in the next section.

The trimer propagation parameter describes the straightening of the third molecule and the formation of steric zipper contacts with the first two molecules (Fig. 1d).

$$-k_BT\ln g_3 = 2f_{\mathrm{SZ}} + f_{\mathrm{CE}} \quad (13)$$

The propagation parameters give the free energy of aggregation per amino acid, however, only half the amino acids in a β-strand participate in steric zipper contacts because the other half remain on the solvent exposed surface. Therefore, f_{SZ} actually represents one-half of the (average) free energy of the steric zipper interaction by a single sidechain. The factor of two in Eq. 13 accounts for the intercalation of the molecule 3's sidechains between molecules 1 and 2 allowing it to form two sets of steric zipper interactions (Fig. 1d). To estimate the value of g_3 we need to know the strength of the steric zipper interactions. This can be obtained from the binding affinity of the fourth molecule.

The fourth molecule can form H-bond contacts with the third molecule while also forming steric zipper interactions with the second molecule in the original dimer

$$-k_B T \ln g_4 = f_{\mathrm{HB}} + f_{\mathrm{SZ}} + f_{\mathrm{CE}} \quad (14)$$

which is the same set of interactions found in mature fibrils. Solubility measurements give ln $g_4 \simeq 0.5$[2,35], so $f_{\mathrm{SZ}} \simeq -0.35 k_B T$ and $-\ln g_3 \simeq 1.06$. Finally, the hairpin monomer propagation parameter is $-k_B T \ln g_{hp} = f_{\mathrm{SZ}} + 2 f_{\mathrm{CE}} \simeq 2.87 k_B T$.

This partitioning of energy reduces the original six parameters to just three; f_{HB}, f_{SZ}, and f_{CE}. Next, we need to know how denaturants will affect these binding energies. To obtain this functionality we make two assumptions. First, we assume that the denaturant will have a linear effect on the binding free energy $f_i(c_d) = f_i(0) + m_i c_d$ where f_i is the negative log of a propagation parameter, c_d is the denaturant concentration, and m_i is a coefficient describing the effect of the denaturant. Secondly, we assume that the denaturant affects the H-bond and steric zipper interactions such that the *m*-value is proportional to the non-entropic contribution to the free energy. This gives

$$\frac{m_s}{f_{\mathrm{HB}}} = \frac{m_2}{f_{\mathrm{HB}}} = \frac{m_3}{2 f_{\mathrm{SZ}}} = \frac{m_4}{f_{\mathrm{HB}} + f_{\mathrm{SZ}}} \quad (15)$$

for helices, dimers, trimers, and mature fibril contacts, respectively. As a rough check of this analysis we calculate the *m*-value for mature fibrils. The urea *m*-value for helices is 0.047 $k_B T \mathrm{M}^{-1}$ [29], so for mature fibrils we expect $m_4 = (f_{\mathrm{HB}} + f_{\mathrm{SZ}}) m_s / f_{\mathrm{HB}} = 0.056\ k_B T \mathrm{M}^{-1}$. We can obtain an estimate for the effect of GdnHCl by noting that the ratio of Gdn and urea m-values for average proteins is 25/13.1[36]. This gives a Gdn m-value of 0.11 $k_B T \mathrm{M}^{-1}$, which is remarkably close to the value 0.12 $k_B T \mathrm{M}^{-1}$ obtained by fitting fibril growth rates[12]. We caution that this analysis, at best, applies to average values and, given the number of assumptions made above, this agreement may very well be a coincidence. However, there is less ambiguity to the main conclusion of this section, which is that fibril must achieve a minimum length of $1 + f_{CE}/\ln g_4$ β-strands to pay the entropic penalty of initiating the fibril. This suggests that the minimum length is 4 or 5 strands, in rough agreement with simulation studies[37–39].

D. Equilibrium dimer and trimer concentrations in solutions of single strand molecules

The propagation constants g_2 and g_3 are both less than unity indicating that the dimer and trimer states are less favorable than disordered monomers. The populations of dimers and trimers are further suppressed by the presence of the favorable helix state and the translational entropy of the monomers. To capture the latter effect we start with the grand free energy of a solution of monomers, dimers, and trimers

$$\mathscr{F}=\sum_{n=1}^{3}[F^{(n)}c_n-nc_n\mu+k_BT(c_n\ln c_n-c_n)] \quad (16)$$

where c_n is the concentration of a species containing n protein molecules. Here $F^{(n)}$ is the free energy of an oligomer containing n molecules, which we obtain from the partition functions calculated above. In the second term the chemical potential μ serves the usual function of a Langrange multiplier to constrain the total protein concentration. The final terms represent the translational entropy of the oligomers. Taking the derivative with respect to c_n we solve for the concentration of each species

$$\frac{\partial F_G}{\partial c_n}=F^{(n)}-n\mu+k_BT\ln c_n=0 \quad (17)$$

which yields

$$c_n=e^{(-F^{(n)}+n\mu)/k_BT} \quad (18)$$

In particular, the expression for c_1 yields an expression for the chemical potential in terms of the monomer concentration

$$\mu=F^{(1)}+k_BT\ln c_1 \quad (19)$$

Thus the dimer and trimer concentrations are

$$c_2=c_1^2e^{-(F^{(2)}-2F^{(1)})/k_BT} \quad (20)$$

$$c_3=c_1^3e^{-(F^{(3)}-3F^{(1)})/k_BT} \quad (21)$$

Since the dimer and trimer are both thermodynamically disfavored, it is an excellent approximation to equate the monomer concentration with the total protein concentration $c_1 \simeq c_t$.

In order to obtain the dimensionless concentrations required by Eqs. 20 and 21, we adopt a lattice gas approximation in which the translational degrees of freedom are discretized by the size of a water molecule. Therefore, the dimensionless concentrations are given by the molarity of a given species divided by 55.5 M, the concentration of pure water. Due to this rough approximation, we do not expect a quantitative agreement between our predictions and experimental concentrations.

Fig. 2 plots the fraction of proteins in the dimer and trimer states as a function of the total protein concentration. The functional form is a simple power law as seen in Eqs. 20 and 21. Since the interaction energies are net repulsive for oligomers of this size, these states roughly correspond to random collisions and are relatively rare until the concentration reaches 10^{-4} M. This concentration, which is approximately 1 mg/ml for the $L = 100$ proteins used in Fig. 2, is the point where the $c_1 \simeq c_t$ approximation begins to break down. At higher concentrations the oligomer concentration can be determined by using Eqs. 20 and 21 to solve the third order polynomial $c_t = c_1 + 2c_2 + 3c_3$ for c_1.

Interestingly, the folded helix state has a relatively small effect on the population of oligomers. For $s = 1.18$ about 86% of the amino acids are in the helical state (Fig. 2a), yet the trimer population is suppressed by less than a factor of 3 and the dimers are only suppressed by about 35%. This finding only applies to equilibrium states; we will soon find that the folded state has a dramatic effect on the kinetics.

E. Nucleation kinetics

We assume that nucleation occurs in a supersaturated solution in which the states occurring before the nucleation barrier have reached a quasi-equilibrium. In the case of single stranded molecules this includes folded and unfolded monomers, dimers, and trimers, while in the case of hairpin molecules it includes monomers in the folded, unfolded and hairpin states. This local equilibrium is possible because of the substantial free energy barrier separating these states from the large aggregates that are the global free energy minimum.

To describe the nucleation time, we modify the rate equation from Classical Nucleation Theory as described previously[16]

$$\begin{aligned} k_{\text{ss}} &= \sum_{m_3=1}^{L} k_{\text{on}}\, c_1 c_3(m_3) \mathscr{E}_1(m_3) \\ k_{\text{hp}} &= \sum_{m_{hp}=1}^{L/2} k_{\text{on}}\, c_1 c_{hp}(m_{hp}) \mathscr{E}_1(2m_{hp}), \end{aligned} \tag{22}$$

Eqs. 22 includes the three ingredients for successful nucleation that are described by Classical Nucleation Theory. First, there needs to be an equilibrium fluctuation large enough to generate the species at the top of the free energy barrier. This is described by the terms $c_3(m_3)$ and $c_{hp}(m_{hp})$, which give the concentration of clusters in solution that present an ordered binding surface of m_3 and $2m_{hp}$ amino acids, respectively. Using Eqs. 9, 10, and 21 these concentrations are given by

$$\begin{gathered}c_3(m_3)=c_1^3\sum_{m_2=m_3}^{L}(m_2-m_3+1)\,(L-m_2+1)^2(L-m_3+1)\left(\frac{g_2}{\lambda_1^2}\right)^{m_2}\left(\frac{g_3}{\lambda_1}\right)^{m_3}\\ c_{hp}(m_{hp})\simeq c_1\sum_{m_{\text{loop}}}(L-2m_{hp}-m_{\text{loop}}+1)\frac{1}{m_{loop}^{3/2}}\left(\frac{g_{hp}}{\lambda_1^2}\right)^{m_{hp}}\end{gathered}\qquad(23)$$

Second, a nucleation attempt begins when an additional molecule binds to the trimer or hairpin causing the cluster to take an initial step downhill in free energy. These attempts are described by the reaction rates $k_{\text{on}}c_1c_3$ or $k_{\text{on}}c_1c_{hp}$. We assume that the rate coefficient k_{on} is limited by the diffusion of the monomers and the probability ϕ that the contact point on the monomer is unfolded (see Eq. 4). Using the rate of reactive particles striking an absorbing sphere, we approximate the collision rate of unfolded molecules as

$$k_{\text{on}}=4\pi aD\phi\qquad(24)$$

where a is the radius of the absorbing surface and D is the diffusion constant of the monomers.

Finally, successful nucleation requires that the newly formed clusters continue to grow without dissolving back to a state below the nucleation barrier. In most pre-nucleation solutions (except in cases of extreme supersaturation), the average time required for a new monomer to diffuse to a growing cluster is longer than the average time it takes for a monomer to detach from the mostly disordered cluster. Therefore, successful nucleation requires a succession of unlikely events where either the diffusion time is shorter than average or the residence time is longer than average so that the cluster experiences net growth. The probability of this happening is given by the factor ε_1 in Eq. 22, which is conceptually identical to the Zeldovich factor in Classical Nucleation Theory[17,40]. In Eqs. 22 ε_1 is written as a function of the number of β-ordered amino acids available for an incoming molecule to bind.

To model the probability of a successful nucleation attempt, we treat the size of the pre-nucleation cluster as a one-dimensional random walk. Forward steps occur when a diffusing monomer binds to the cluster causing it to grow. This occurs with a rate c_1k_{on}. Reverse steps happen when a molecule detaches from the cluster. For this to occur, the molecule must break all of the H-bonds holding it to the cluster. If the cluster is highly ordered, the molecules can form more bonds and it takes longer before they are all broken at the same time. As a rough approximation we might expect the residence time of a bound molecule to have a simple Arrhenius dependence $\tau_{\text{res}}\sim g_4^m$, where m is the number of H-bonds to be broken ($m = m_3$ or $2m_3$ for the single strand and hairpin cases, respectively)[41]. A more careful calculation gives[12]

$$t_{\mathrm{res}}=-\frac{1}{v_b}+\frac{D_b}{v_b^2}e^{v_b m/D_b}(1-e^{-v_b/D_b}). \qquad (25)$$

where v_b and D_b are the effective drift velocity and diffusion constant of the reaction coordinate describing the number of H-bonds. These are given by

$$v_b=k_+-k_-=k_+(1-g_4^{-1}) \qquad (26)$$

$$D_b=\frac{1}{2}(k_++k_-)=\frac{k_+}{2}(1+g_4^{-1}) \qquad (27)$$

where $k_+ \simeq \mathrm{ns}^{-1}$ is the timescale for the formation of an H-bond [42] and we have used detailed balance to relate the rates of H-bond formation and breakage (k_-) to the free energy of the bonds, $k_+/k_- = g_4$.

With Eqs. 24 and 25 we can determine the probability that the cluster gains or loses a molecule

$$\begin{aligned} p_+ &= \frac{c_1 k_{on}}{c_1 k_{on}+k_{res}} \\ p_- &= \frac{k_{res}}{c_1 k_{on}+k_{res}} \end{aligned} \qquad (28)$$

where the rate of molecular detachment is $k_{\mathrm{res}}=t_{\mathrm{res}}^{-1}$. Equations 28 define two concentration regimes for nucleation. When $c_1 k_{\mathrm{on}} > k_{\mathrm{res}}$ new molecules generally add to the cluster faster than they fall off. This means that the rate limiting step for nucleation is the formation of the state at the top of the free energy barrier, since this state has a high probability of continuing to grow. On the other hand, at physiological concentrations we expect that the opposite limit $c_1 k_{\mathrm{on}} < k_{\mathrm{res}}$ holds. In this regime the cluster is more likely to lose molecules than add them. Therefore, nucleation requires the unlikely event where many molecules add with few detachments. In other words, the cluster size performs a random walk that is biased toward shrinkage events. We define a nucleation attempt to begin when a cluster grows larger than the most unstable size n_c. The attempt fails when the cluster returns to n_c and succeeds when it reaches the stable size N^*. Therefore, the success probability ε_1 is the probability of a walk that starts at $n_c + 1$ and reaches N^* without returning to n_c. If N is the size of the cluster, a convenient reaction coordinate is $n = N - n_c$, the number of molecules above the most unstable size, where $n_c = 1$ or 3 for the hairpin and single strand cases, respectively.

The nucleation probability ε_1 probability satisfies the recursion relation[43]

$$\mathscr{E}_n = p_+ \mathscr{E}_{n+1} + p_- \mathscr{E}_{n-1}, \tag{29}$$

reflecting the fact that a cluster with n molecules evolves to a cluster with $n + 1$ molecules with probability p_+ or at $n - 1$ with probability p_-. Eq. 29 can be rewritten as the matrix equation $\boldsymbol{\mu}(n + 1) = \boldsymbol{M}\boldsymbol{\mu}(n)$ where

$$\boldsymbol{u}(n) = \begin{pmatrix} \mathscr{E}_n \\ \mathscr{E}_{n-1} \end{pmatrix} \tag{30}$$

$$\boldsymbol{M} = \begin{pmatrix} \frac{1}{p_+} & 1 - \frac{1}{p_+} \\ 1 & 0 \end{pmatrix}. \tag{31}$$

The transfer matrix can be brought into diagonal form with the transformation

$$\boldsymbol{T}^{-1}\boldsymbol{M}\boldsymbol{T} = \begin{pmatrix} \frac{1}{\frac{1}{p_+} - 2} & \frac{-1}{\frac{1}{p_+} - 2} \\ \frac{-1}{\frac{1}{p_+} - 2} & \frac{\frac{1}{p_+} - 1}{\frac{1}{p_+} - 2} \end{pmatrix} \begin{pmatrix} \frac{1}{p_+} & 1 - \frac{1}{p_+}(14) \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{p_+} - 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{32}$$

$$= \begin{pmatrix} \frac{1}{p_+} - 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{33}$$

By applying the transfer matrix we can generate the success probability for a cluster of any size

$$\boldsymbol{u}(n) = \boldsymbol{M}^{n-1}\boldsymbol{u}(1) \tag{34}$$

$$= \boldsymbol{T}\left(\boldsymbol{T}^{-1}\boldsymbol{M}\boldsymbol{T}\right)^{n-1}\boldsymbol{T}^{-1}\boldsymbol{u}(1) \tag{35}$$

$$=\begin{pmatrix} \mathscr{E}_1\frac{\left(\frac{1}{p_+}-1\right)^n-1}{\frac{1}{p_+}-2} \\ \mathscr{E}_1\frac{\left(\frac{1}{p_+}-1\right)^{n-1}-1}{\frac{1}{p_+}-2} \end{pmatrix} \quad (36)$$

where we have used the boundary condition $\mathscr{E}_0 = 0$. By applying the other boundary condition $\mathscr{E}_{n^*} = 1$ we arrive at the desired result

$$\mathscr{E}_1=\frac{\frac{1}{p_+}-2}{\left(\frac{1}{p_+}-1\right)^{N^\star-n_c}-1} \quad (37)$$

Fig. 3 shows how the nucleation success probability depends on the monomer concentration and the residence time of the bound molecules. At low concentration the detachment rate greatly exceeds the rate of new molecules resulting in prohibitively low success probabilities. The probability increases when either the concentration increases or the nucleating cluster becomes more ordered which increases the residence time of newly bound molecules.

The final result for the nucleation rate is given by Eq. 22 with Eqs. 23, 24, and 37.

III. RESULTS AND DISCUSSION

A. Effect of folded state on nucleation

The nucleation rate predicted by Eq. 22 is plotted in Fig. 4 for molecules with ($s \neq 0$) and without ($s = 0$) a stable folded state. In the absence of denaturant the folded state suppresses nucleation by 7 orders of magnitude. Addition of denaturant leads to a rapid increase in the nucleation of the folded protein because the folding equilibrium shifts toward the aggregation-prone unfolded state. However, the denaturant also destabilizes the aggregated state, as seen by the declining nucleation rate of the intrinsically disordered molecules. As a result of these competing effects, the nucleation rate reaches a maximum at about 3 M GdnHCl. At this point the monomeric protein is mostly unfolded and the nucleation rate for the folded and intrinsically disordered cases converge. Urea has a weaker denaturing effect and does not reach a maximum nucleation rate until the concentration is above 6 M. We note that the two order of magnitude enhancement between 2 M and 4 M urea (Fig. 4) is qualitatively consistent with the observation that the lag time for lysozyme aggregation disappears over this range[44].

B. Off-register binding

Molecular models of mature fibrils show a striking level of order[13–15]. Most commonly, the molecules form parallel, in-register β-sheets, although anti-parallel structures have also been observed[45,46]. It is an open question whether this perfect alignment of sidechains is representative of all fibrils or simply an artifact of structural techniques that are most sensitive to ordered structures. From a self-assembly perspective, we expect that slow growth conditions will favor more ordered structures while rapid growth conditions will promote the incorporation of defects[12,47]. This would suggest that natural fibrils grown at physiological concentrations would tend to be more ordered (provided only one protein species is incorporated) while the higher concentrations employed *in vitro* would lead to more disorder.

The nucleation model presented above ignores sequence effects in that all binding states are treated as equivalent. This is the relevant case when considering the aggregation of homopolymers like polyglutamine or very high supersaturations where disordered aggregates are expected to grow. In addition, if the binding selectivity is enforced by steric complementarity more than the chemistry of the sidechains, the small size of the pre-nucleation cluster may allow enough conformational lability to permit promiscuous binding[48]. This would allow for a two-step nucleation process in which cluster formation precedes the onset of crystal order[22,49–51]. A similar decoupling of the density and alignment order parameters is thought to be the nucleation mechanism in protein crystals [52,53].

The opposite limit, where the binding registry is strictly enforced, will modify the theory in two ways. First, it will sharply reduce the concentration of pre-nucleation clusters due to the reduced degeneracy of binding. Eqs. 23 then become

$$\begin{gathered} c_3(m_3) \simeq c_1^3 \sum_{m_2=m_3}^{L} (m_2 - m_3 + 1)\,(L - m_2 + 1) \left(\frac{g_2}{\lambda_1^2}\right)^{m_2} \left(\frac{g_3}{\lambda_1}\right)^{m_3} \\ c_{hp}(m_{hp}) = \sum_{m_{\text{loop}}} \frac{1}{m_{loop}^{3/2}} \left(\frac{g_{hp}}{\lambda_1^2}\right)^{m_{hp}} \end{gathered} \tag{38}$$

The single strand expression has a degeneracy factor describing the choice of m_2 residues out of L for the location of the H-bonds and a second factor to describe where the third molecule inserts its sidechains to form the steric zipper contacts. The hairpin structure, on the other hand, is uniquely determined by the length of the steric zipper interface, m_{hp} and the size of the disordered loop.

Secondly, in-register binding will occur at a much lower rate than off-register binding. If there are L amino acids in each protein, we expect that in-register binding will occur with a probability L^{-1}. This is equivalent to increasing the diffusion time by a factor of L. This has a large effect on ε_1, which scales with the diffusion according to $(c_1 k_{\text{on}})^{2-N^*}$. As a result, the requirement of in-register binding reduces the nucleation rate greatly relative to the promiscuous binding case (Fig. 5).

C. Heterogeneous nucleation

Solution impurities can increase nucleation rates by providing binding surfaces for the particles. The energy of binding to the impurity partially offsets the entropic penalty of bringing the particles together, thereby increasing the concentration of critical species. In amyloid systems the nucleation barrier is due to both the translational entropy cost of creating a high density fluctuation and the conformational entropy of stretching the proteins into β-strand conformation. Therefore, heterogeneous binding sites that favor elongated molecules can provide a particularly advantageous pathway to nucleation. A favorable conformational bias can be provided by a surface that is planar on the length scale of the β-strands. Such surfaces include membranes, air-water interfaces, oil droplets, or even the sides of existing fibrils.

To model heterogeneous nucleation we compute the concentration of an assembly of n molecules bound to a heterogeneous site

$$c_{Hn}(m_3) = c_{\mathrm{het}} c_1^n e^{-f_{\mathrm{Hn}}(n)/k_B T} \qquad (39)$$

where c_{het} is the concentration of heterogeneous binding sites and f_{Hn} is the free energy of the binding site-oligomer complex. Generalizing Eq. 23 for single stranded trimers we have

$$c_{H3}(m_3) \simeq c_{\mathrm{het}} c_1^3 \sum_{m_2=m_3}^{L} (m_2-m_3)(L-m_2)^2 (L-m_3) \left(\frac{g_2}{\lambda_1^2}\right)^{m_2} \left(\frac{g_3}{\lambda_1}\right)^{m_3} e^{-m_{\mathrm{surf}} f_{\mathrm{het}/k_B T}} \qquad (40)$$

Here f_{het} is the binding energy per amino acid between the proteins and the heterogeneous surface and m_{surf} is the number of amino acids bound to the surface. The binding energy f_{het} depends strongly on the nature of the heterogeneous binding site with inert surfaces contributing zero binding energy. Depending on whether the protein-surface interaction occurs via sidechains or backbone H-bonds, m_{surf} can be either $2m_2$ or $m_2 + m_3$. Here we assume the protein-impurity interaction is mediated by sidechains so $m_{\mathrm{surf}} = 2m_2$. We have also made the assumption that the allowed Ramachandran space is sufficiently limited that binding to a planar surface also restricts the molecules to conformations closely approximating β-strands.

Heterogeneous nucleation will dominate the system when $c_{H3} > c_3$, therefore, the required concentration of impurity sites for heterogeneous nucleation to be significant is $c_3 c_{\mathrm{het}}/c_{H3}$. This quantity is plotted in Fig. 6 as a function of the impurity binding energy. The surface binding energy has an exponential effect on the trimer concentration with a marked change in the exponent at $f_{\mathrm{het}} \simeq 1\ k_B T$. This value corresponds to the point where free energy of forming the trimer switches from net unfavorable to favorable. When this happens the partition function for the trimer states becomes dominated by the highly ordered terms, leading to the abrupt change in the slope in Fig. 6.

A particularly important case of heterogenous nucleation is that of secondary nucleation, where existing fibrils provide the substrate for nucleation events[57]. A recent simulation study showed that Aβ monomers form favorable interactions with hydrophobic sidechains on the fibril surface causing them to extend parallel to the fibril axis[58]. These sidechain mediated interactions are qualitatively similar to steric zipper interactions, yet presumably stronger since the monomer will favor the most attractive sidechains on the fibril surface. This suggests a heterogeneous binding energy on the order of $f_{\text{het}} \simeq 2f_{\text{CE}} \simeq 1.6\ k_BT$.

IV. CONCLUSION

We have presented a toy model for the nucleation of amyloid fibrils from proteins that have a stable folded state. Experiments have shown that the fibril state is much more stable than the natively folded state, so the folded state represents a deep kinetic trap that helps prevent aggregation[35]. Our calculations show that the native state has a profound effect on nucleation kinetics (Fig. 4) but only a modest suppression on the concentration of unstable oligomers that provide the substrate for nucleation (Fig. 2b). This explains why destabilizing factors, like increased temperature or the addition of denaturants, often lead to rapid aggregation.

Due to the difficulty in disentangling the effects of secondary nucleation and fragmentation from primary nucleation, direct measurements of the primary nucleation rate are sparse. Figure 5 shows three such measurements along with the predictions from our theory. The most direct measurement of the nucleation rate used insulin in micron-scale droplets coated with surfactants to eliminate heterogeneous nucleation[56]. This setup allowed individual nucleation event to be directly resolved, thereby eliminating the complicating factor of secondary nucleation. These experiments yielded a nucleation rate of $5.6\times10^6\text{s}^{-1}\text{L}^{-1}$ at 6 mM protein concentration. This rate is in good agreement with the predicted rate for single stranded molecules, which is surprising because a molecule of the size of insulin would be expected to nucleate using the much faster hairpin mechanism. Unfortunately, the concentration dependence was not investigated in these experiments. This shortcoming was addressed in later studies using $\text{A}\beta_{40}$ and $\text{A}\beta_{42}$ as test systems[54,55]. These works extracted the nucleation rate coefficient by fitting the time dependent fibril concentration to a kinetic theory accounting for secondary nucleation. The obtained rates lie between our predictions for single stranded and hairpin molecules. Assuming that $\text{A}\beta$ nucleates via the hairpin mechanism, the experiments are in much better agreement with the theory that assumes that the amino acid alignment is strictly enforced. However, the numerical discrepancy grows to nearly ten orders of magnitude at the upper end of the experimental concentration range. While this is a large number, it is comparable to the discrepancy found in applying nucleation theory to other protein systems[59]. In addition, the strong exponential and power law dependencies inherent to nucleation ensure that small errors are greatly magnified and rough approximations, such as our scaling of concentration units, could be contributing here.

More useful information can be obtained from the concentration dependence. Our theory predicts two power law regimes for the concentration dependence. At high concentrations, where $ck_{\text{on}} > k_{\text{res}}$, new molecules bind to the cluster faster than they detach, meaning that the success probability ε_1 saturates near unity (Fig. 3). Therefore, the concentration dependence comes from the concentration of unstable clusters and the diffusion rate. This gives a concentration dependence of c^{n_c+1}, which results in c^2 for hairpin molecules. While this power law agrees with the experimental data for both $\text{A}\beta$ systems [54,55], it is surprising that this limit applies to the concentrations where the experiments were conducted. At the μM concentrations explored, the diffusion rates are on the order of $ck_{\text{on}} \simeq 10^3\ \text{s}^{-1}$. To achieve a detachment rate slower than this would require an ordered binding surface of ~ 25 amino acids (Fig. 3). Since the entropic cost of ordering amino acids is nearly 2 k_BT, we would expect that the system is in the low concentration limit, $ck_{\mathrm{on}} < k_{\mathrm{res}}$. In this limit the success probability varies with concentration like c^{N^*}, meaning that the overall nucleation rate varies like c^{N^*}. The resulting prediction of a nucleation rate proportional to c^5 can be ruled out by current experiments (Fig. 5).

How is it that the system is actually in the high concentration regime? The most likely explanation is that nucleation is occurring by a heterogeneous mechanism. Binding to impurity sites will shift the free energy landscape, but will not alter the overall scaling behavior. In this case, binding to an impurity could align the initial molecule enough that the binding surface for subsequent molecules exceeds the 25 amino acids estimated above. We do not believe that a more complicated pathway, for example a two-step mechanism, could explain the observed weak power law because even the smallest disordered cluster, a dimer, would bring a concentration dependence of c^2 with subsequent addition events bringing additional powers.

An important caveat is that the nucleation theory presented here uses a one-dimensional reaction coordinate (N). This means that it is unable to capture the displacement of the nucleation flux away from the free energy saddle point[16]. The summation over core sizes in Eq. 22 has a peak flux for clusters with ordered cores of $m \simeq 3$ at all concentrations. However, intuition suggests that the flux should shift to larger cores at lower concentrations. This is because the increased waiting time will give the system more time to explore ordered states with high free energy. This could also contribute to the overly strong concentration dependence predicted by the model and the discrepancy in the magnitude of the rates.

Another limitation of our model is that the helix-coil model lacks the cooperativity found in proteins with more complicated folds[29]. This, coupled with the rough estimates used in our parameters, means that our predictions are unlikely to be quantitatively accurate. Still, our simple model provides needed insight into the energetics and scaling behavior of fibril nucleation.

Acknowledgments

This work was supported by NIH Grant R01GM107487.

References

1. Lee CF. Phys Rev E. 2009; 80:31922.
2. Schmit JD, Ghosh K, Dill KA. Biophys J. 2011; 100:450–8. [PubMed: 21244841]
3. Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, Vendruscolo M. J Mol Biol. 2008; 380:425–36. [PubMed: 18514226]
4. Dill KA, Ghosh K, Schmit JD. Proc Natl Acad Sci U S A. 2011; 108:17876–82. [PubMed: 22006304]
5. Prusiner SB. Proc Natl Acad Sci. 1998; 95:13363–13383. [PubMed: 9811807]
6. Miti T, Mulaj M, Schmit JD, Muschol M. Biomacromolecules. 2015; 16:326–335. [PubMed: 25469942]

7. Knowles TPJ, Waudby Ca, Devlin GL, Cohen SIa, Aguzzi A, Vendruscolo M, Terentjev EM, Welland ME, Dobson CM. Science. 2009; 326:1533–7. [PubMed: 20007899]

8. Cohen, SIa, Vendruscolo, M., Welland, ME., Dobson, CM., Terentjev, EM., Knowles, TPJ. J Chem Phys. 2011; 135:65105.

9. Cohen, SIa, Vendruscolo, M., Dobson, CM., Knowles, TPJ. J Chem Phys. 2011; 135:65106.

10. Cohen, SIa, Vendruscolo, M., Dobson, CM., Knowles, TPJ. J Chem Phys. 2011; 135:65107.

11. Lee CF, Loken J, Jean L, Vaux DJ. Phys Rev E. 2009; 80:41906.

12. Schmit JD. J Chem Phys. 2013; 138:185102. [PubMed: 23676074]

13. Lührs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Döbeli H, Schubert D, Riek R. Proc Natl Acad Sci U S A. 2005; 102:17342–7. [PubMed: 16293696]

14. Wasmer C, Lange A, Van Melckebeke H, Siemer AB, Riek R, Meier BH. Science. 2008; 319:1523–6. [PubMed: 18339938]

15. Nelson R, Sawaya MR, Balbirnie M, Madsen AØ, Riekel C, Grothe R, Eisenberg D. Nature. 2005; 435:773–8. [PubMed: 15944695]

16. Zhang L, Schmit JD. Phys Rev E. 2016; 93:60401.

17. Kashchiev, D. Nucleation. Butterworth-Heinemann; 2000.

18. Zhang J, Muthukumar M. J Chem Phys. 2009; 130:35102.

19. Cabriolu R, Kashchiev D, Auer S. J Chem Phys. 2010; 133:225101. [PubMed: 21171698]

20. Auer S. J Phys Chem B. 2014; 118:5289–99. [PubMed: 24784223]

21. Auer S. Biophys J. 2015; 108:1176–1186. [PubMed: 25762329]

22. Serio TR, Cashikar AG, Kowal AS, Sawicki GJ, Moslehi JJ, Serpell L, Arnsdorf MF, Lindquist SL. Science (80-). 2000; 289:1317–1321.

23. Hill SE, Robinson J, Matthews G, Muschol M. Biophys J. 2009; 96:3781–90. [PubMed: 19413984]

24. Seubert P, Vigo-Pelfrey C, Esch F, Lee M, Dovey H, Davis D, Sinha S, Schlossmacher M, Whaley J, Swindlehurst C. Nature. 1992; 359:325–7. [PubMed: 1406936]

25. Sengupta P, Garai K, Sahoo B, Shi Y, Callaway DJE, Maiti S. Biochemistry. 2003; 42:10506–13. [PubMed: 12950178]

26. Laganowsky A, Liu C, Sawaya MR, Whitelegge JP, Park J, Zhao M, Pensalfini A, Soriaga AB, Landau M, Teng PK, et al. Science. 2012; 335:1228–1231. [PubMed: 22403391]

27. Apostol MI, Perry K, Surewicz WK. J Am Chem Soc. 2013; 135:10202–5. [PubMed: 23808589]

28. Nagel-Steger L, Owen MC, Strodel B. Chem Bio Chem. 2016; 17:657–676.

29. Ghosh K, Dill KA. J Am Chem Soc. 2009; 131:2306–12. [PubMed: 19170581]

30. Zimm BH, Bragg JK. J Chem Phys. 1959; 31:526.

31. Poland, D., Scheraga, HA. Theory of Helix-Coil Transitions in Biopolymers: Statistical Mechanical Theory of Order-Disorder Transitions in Biological Macromolecules. Academic Press; 1970.

32. Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, Delaglio F, Tycko R. Proc Natl Acad Sci U S A. 2002; 99:16742–7. [PubMed: 12481027]

33. Luca S, Yau WM, Leapman R, Tycko R. Biochemistry. 2007; 46:13505–13522. [PubMed: 17979302]

34. Doi, M., Edwards, SF. The Theory of Polymer Dynamics (International Series of Monographs on Physics). Oxford University Press; Oxford: 1988.

35. Baldwin AJ, Knowles TPJ, Tartaglia GG, Fitzpatrick AW, Devlin GL, Shammas SL, Waudby Ca, Mossuto MF, Meehan S, Gras SL, et al. J Am Chem Soc. 2011; 133:14160–3. [PubMed: 21650202]

36. Ghosh K, Dill KA. Proc Natl Acad Sci U S A. 2009; 106:10649–54. [PubMed: 19541647]

37. Zheng J, Ma B, Tsai CJ, Nussinov R. Biophys J. 2006; 91:824–833. [PubMed: 16679374]

38. De Simone A, Esposito L, Pedone C, Vitagliano L. Biophys J. 2008; 95:1965–1973. [PubMed: 18469082]

39. Kahler A, Sticht H, Horn AHC. PLoS One. 2013; 8doi: 10.1371/journal.pone.0070521

40. Zeldovich YB. Acta Physicochim URSS. 1943; 18:1–22.

41. Schmit J. n.d
42. Muñoz V, Thompson Pa, Hofrichter J, Eaton Wa. Nature. 1997; 390:196–9. [PubMed: 9367160]
43. Redner, S. A Guide to First-Passage Processes. Cambridge University Press; 2007.
44. Kumar EK, Prabhu NP. Phys Chem Chem Phys. 2014; 16:24076–24088. [PubMed: 25288276]
45. Balbach JJ, Ishii Y, Antzutkin ON, Leapman RD, Rizzo NW, Dyda F, Reed J, Tycko R. Biochemistry. 2000; 39:13748–59. [PubMed: 11076514]
46. Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers Sa, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJW, McFarlane HT, et al. Nature. 2007; 447:453–7. [PubMed: 17468747]
47. Whitelam S, Schulman R, Hedges L. Phys Rev Lett. 2012; 109:265506. [PubMed: 23368583]
48. Cukalevski R, Yang X, Meisl G, Weininger U, Bernfur K, Frohm B, Knowles TPJ, Linse S. Chem Sci. 2015; 6:4215–4233.
49. Auer S, Meersman F, Dobson CM, Vendruscolo M. PLoS Comput Biol. 2008; 4:e1000222. [PubMed: 19008938]
50. Auer S, Dobson CM, Vendruscolo M, Maritan A. Phys Rev Lett. 2008; 101:17–20.
51. Auer S, Ricchiuto P, Kashchiev D. J Mol Biol. 2012; 422:723–30. [PubMed: 22721952]
52. Whitelam S. J Chem Phys. 2010; 132:194901. [PubMed: 20499986]
53. Vekilov PG. Cryst Growth Des. 2010; 10:5007–5019.
54. Cohen, SIa, Linse, S., Luheshi, LM., Hellstrand, E., White, Da, Rajah, L., Otzen, DE., Vendruscolo, M., Dobson, CM., Knowles, TPJ. Proc Natl Acad Sci U S A. 2013; 110:9758–63. [PubMed: 23703910]
55. Meisl G, Yang X, Hellstrand E, Frohm B, Kirkegaard JB, Cohen SIa, Dobson CM, Linse S, Knowles TPJ. Proc Natl Acad Sci. 2014; 111:9384–9389. [PubMed: 24938782]
56. Knowles TPJ, White Da, Abate AR, Agresti JJ, Cohen SIa, Sperling Ra, De Genst EJ, Dobson CM, Weitz Da. Proc Natl Acad Sci U S A. 2011; 108:14746–51. [PubMed: 21876182]
57. Rubio, Ma, Schlamadinger, DE., White, EM., Miranker, AD. Biochemistry. 2015; 54:987–993. [PubMed: 25541905]
58. Barz B, Strodel B. Chem - A Eur J. 2016; 22:8768–8772.
59. Sear RP. J Phys Condens Matter. 2007; 19:33101.

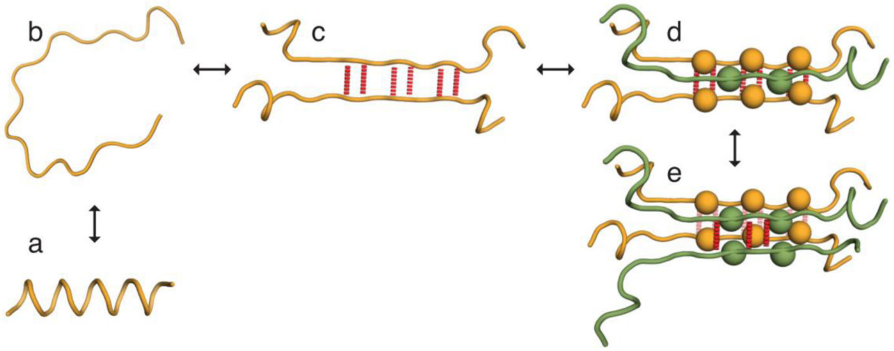

Figure 1.

Cartoon of the aggregation states modeled in our theory. The left panels show the conversion of monomers between (a) an aggregation-resistant folded state that we model as a helix and (b) an aggregation-prone disordered state. (c) A pair of (partially) unfolded proteins forming a dimer through the formation of intermolecular H-bonds. In the figure the dimer is held together by $m_2 = 6$ H-bonds (red dotted lines). (d) A trimer which is held together by H-bonds between the first two molecules (rear) and steric zipper interactions with the third molecule (front). Sidechains participating in the steric zipper are shown as spheres with the remaining sidechains omitted for clarity. The total number of ordered amino acids is $2m_2 + m_3$, m_2 from each of the rear molecules plus m_3 from the front molecule. In the illustrated conformation there are two sidechains from the front molecule participating in the steric zipper. Since every other sidechain is on the opposite face of the β-strand, this requires at least $m_3 = 3$ amino acids to be in the β-sheet conformation. (e) A tetramer with $m_2 = 6$ β-ordered amino acids on each of the molecules in the rear β-sheet (six yellow H-bonds) and three β ordered amino acids on each strand of the front β-sheet (three red H-bonds).

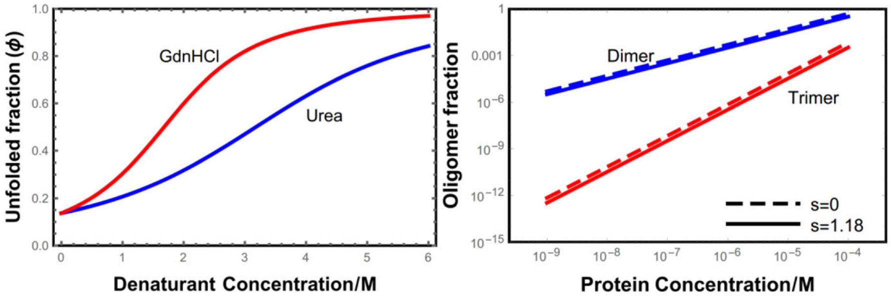

Figure 2.

(a) Denaturant unfolding curves for isolated helices. For the parameters given in the text, helices are ~86% folded in the absence of denaturant. We expect that real proteins will show a more cooperative transition than the helix coil model meaning that they are more folded in the absence of denaturant and the unfolding transition will be more abrupt[29]. (b) Fraction of $L = 100$ helical proteins in the dimer state ($2c_2/c_t$) and in the trimer state ($3c_3/c_t$) as a function of the total protein concentration.

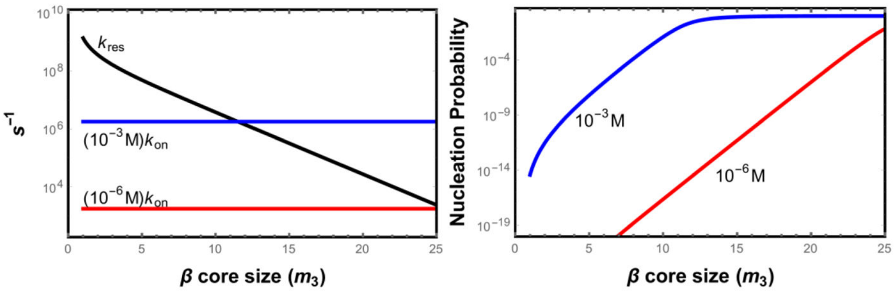

Figure 3.

(a) Comparison of the monomer detachment rate (blue) to the diffusion limited arrival of monomers to the growing nucleus. (b) Probability that a nascent nucleus reaches the stable size as a function of the size of the β core. When the core becomes large enough that newly bound molecules have a residence time comparable to the diffusion time the success probability approaches unity. Both plots assume disordered monomers ($s = 0$).

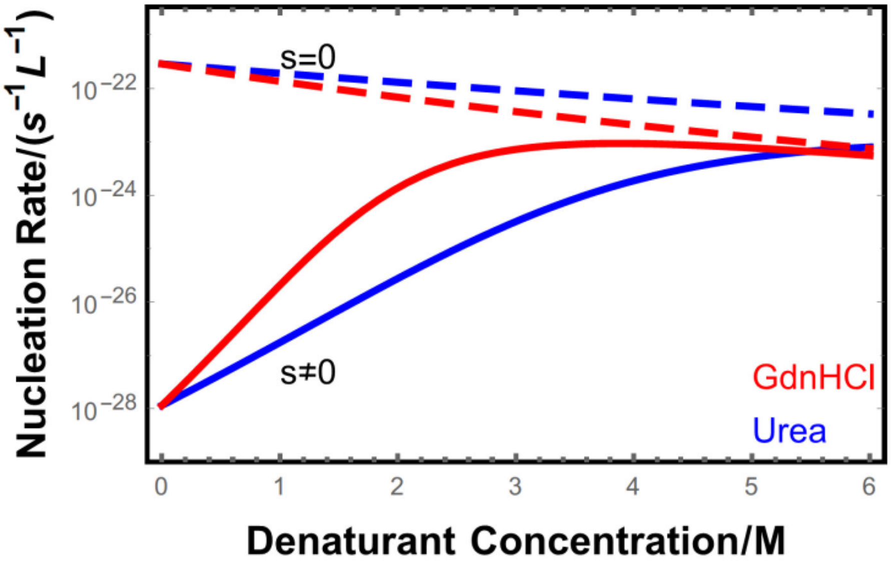

Figure 4.

Nucleation rate as a function of denaturant concentration for proteins of length $L = 100$ at a concentration $c_t = 10^{-5}$ M. Intrinsically disordered proteins ($s = 0$) are monotonically inhibited from nucleating by the addition of denaturant due to the weakening of intermolecular bonds. In contrast, proteins with stable folded states (solid lines) show greatly enhanced nucleation upon denaturant addition because of the increased population of unfolded proteins ($s = 1.16$ at $c_d = 0$). This trend reverses at high denaturant concentrations since the proteins are mostly unfolded and the dominant effect of further denaturant addition is the destabilization of fibril contacts.

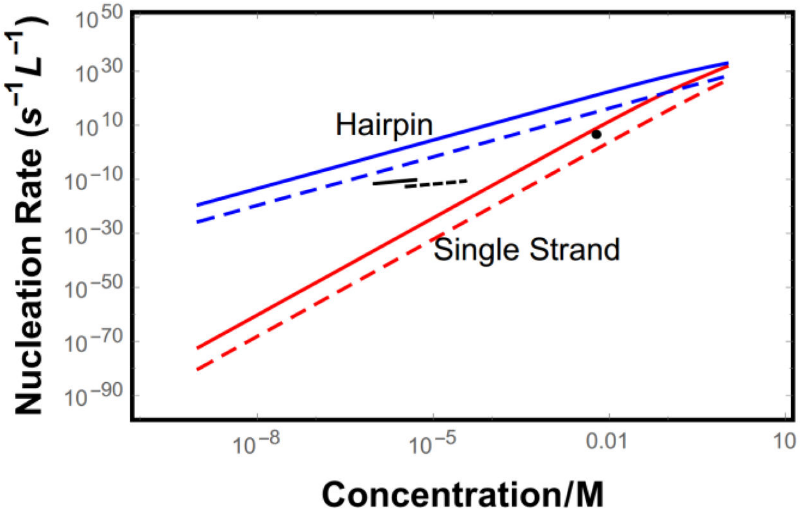

Figure 5.

Predicted nucleation rate for hairpin molecules with L = 40 (blue) and single strand molecules with L = 15 (red). Long dashes represent the case where binding registry is strictly enforced, while solid lines show promiscuous binding. Short black lines show experimentally determined nucleation rates (data were used to fit the proportionality constant and exponent for $k_{nuc} \propto c^x$) for Aβ_{42} (solid)[54] and Aβ_{40} (dashed)[55]. The single black dot shows the nucleation rate for insulin[56].

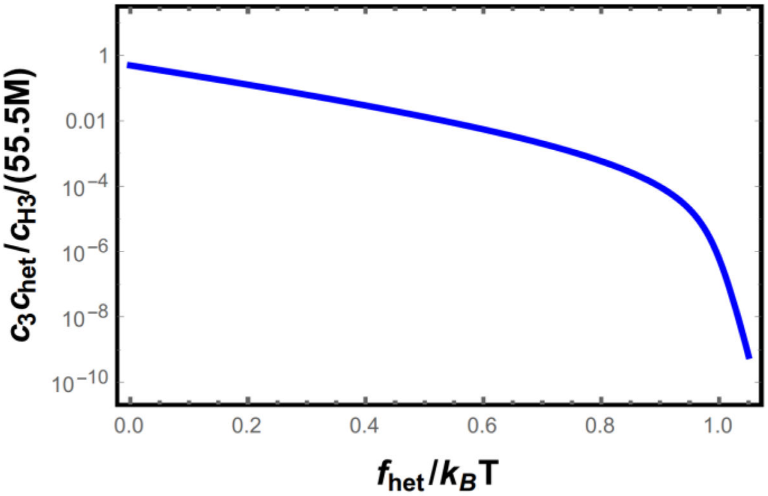

Figure 6.

Required concentration of impurity binding sites needed for heterogeneous nucleation to dominate homogeneous nucleation as the primary nucleation mechanism. There is an abrupt change in the slope when the impurity binding energy, f_{het}, becomes strong enough to overcome the entropic penalty of straightening the proteins.