Manuscript submitted to **Biophysical** *Journal*
Article

Conformational entropy limits the transition from nucleation to elongation in amyloid aggregation

Tien M. Phan[1] and and Jeremy D. Schmit[1*]

[1]Department of Physics, Kansas State University, Manhattan, KS 66506, USA
[*]Correspondence: schmit@phys.ksu.edu

ABSTRACT The formation of β-sheet rich amyloid fibrils in Alzheimer's disease and other neurodegenerative disorders is limited by a slow nucleation event. To understand the initial formation of β-sheets from disordered peptides, we used all-atom simulations to parameterize a lattice model that treats each amino acid as a binary variable with β and non-β states. We show that translational and conformational entropy give the nascent β-sheet an anisotropic surface tension which can be used to describe the nucleus with two-dimensional Classical Nucleation Theory. Since translational entropy depends on concentration, the aspect ratio of the critical β-sheet changes with protein concentration. Our model explains the transition from the nucleation phase to elongation as the point where the β-sheet core becomes large enough to overcome the conformational entropy cost to straighten the terminal molecule. At this point the β-strands in the nucleus spontaneously elongate, which results in a larger binding surface to capture new molecules. These results suggest that nucleation is relatively insensitive to sequence differences in co-aggregation experiments because the nucleus only involves a small portion of the peptide.

SIGNIFICANCE The conversion of soluble proteins to amyloid aggregates is associated with many neurodegenerative diseases. Experiments have shown that this conversion occurs by a slow nucleation step followed by rapid growth. This work identifies the principle contributions to the free energy barrier that separates these two stages. It also shows how factors like protein concentration, sidechain interactions, and interactions with the environment can modify the barrier and affect nucleation times.

INTRODUCTION

The assembly of proteins into amyloid fibrils causes numerous neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, and prion disorders (1). *In vitro* experiments show that the conversion of soluble peptides to the fibril state is limited by nucleation events, indicative of a free energy barrier to initiate the fibril state. The linear morphology of amyloid fibrils suggests that the nucleation barrier must be substantially different than other nucleating systems. In the case of isotropic particle condensation, the nucleation barrier arises because particles at the periphery have sacrificed the translational entropy of the dilute phase, but only form a fraction of the favorable interactions available to interior particles (2). This interaction deficit, usually described as a surface tension, becomes a smaller fraction of the free energy as the cluster grows larger.

Surface tension does not limit 1D assemblies because the surface energy does not depend on the cluster size (3). Early attempts to explain amyloid nucleation identified β-sheet layering as a second assembly dimension (4–9). An alternative mechanism that can result in nucleation behavior in 1D systems is a conformational conversion to the β-sheet state (10). This mechanism has been implicitly included in coarse-grained simulations by modeling the molecules with two states; a high energy but strongly attractive fibrillar state, and a low energy state with weaker interactions (11–13). Due to the large energy difference between these states, it was found that non-fibrillar oligomers were necessary to overcome the energy barrier to initiate a fibril. In this paper we remove the two-state assumption by allowing the molecule to partially convert to the fibril state. This partial conversion enables pathways with a lower barrier to conformational conversion. We find that the nucleation of a β-sheet, and hence an amyloid fibril, is limited by two contributions to the free energy barrier, one arising from the translational entropy cost to recruit molecules from solution, and one arising from the conformational entropy cost to form β-strands. These contributions can be modeled using Classical Nucleation Theory (CNT) with an anisotropic surface tension. The distinct entropic contributions contributing to the surface tension in different directions means that the *size and shape*

of the critical β-sheet is sensitive to factors like the amino acid sequence and protein concentration. We also find that there is a critical fibril length at which the conformational entropy barrier vanishes. At this point the β-strands in the cluster will spontaneously elongate. These longer β-strands provide a stronger binding surface for incoming molecules, which we attribute to the end of the nucleation phase and the beginning of the elongation phase.

METHOD

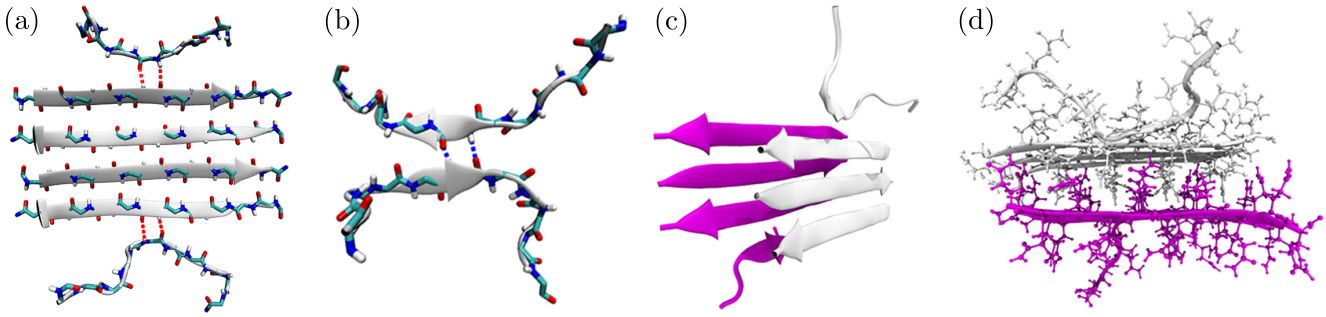

Figure 1: MD snapshots from the sampling of the conversion between β and non-β states. (a) Kinetic parameters for post-nucleation ("strong") bonds are sampled from the terminal molecules on an established cluster (side chains are not shown for clarity). (b) Pre-nucleation ("weak") bonds are sampled using a dimer that is harmonically restrained at the central amino acids. (c) Ribbon representation of the β-sheet bilayer simulation showing the stacking of the two sheets. The six core molecules are restrained in β conformation. The top molecule in the white sheet and the bottom molecule in the magenta sheet are free to sample β and non-β states. This side view shows the staggered alignment of β-strands in which the white strands are positioned between the magenta strands in the opposite sheet. This staggering ensures that each molecule addition results in an equivalent set of H-bond and sidechain interactions. (d) Atomistic view of the bilayer showing the interdigitation of side chains between the two β-sheets.

We employ a dual resolution model to simulate the conversion of peptide strands from the disordered state to a β-sheet. Our approach follows from recent work in which fibril elongation was modeled by treating each amino acid as a binary variable that can be in either β or non-β states (14, 15). We use all-atom molecular dynamics (MD) on a hexamer and dimer of polyglutamine, motivated by the aggregation-prone region of huntingtin protein (16–23), to measure kinetic transitions between the β and non-β states in the post- and pre-nucleation scenarios, respectively (Fig. 1).

Molecular dynamics simulation details

To simulate post-nucleation bonds in the all-atom model we used a β-sheet consisting of six monomers. The backbone atoms of the four internal molecules were harmonically restrained to mimic the rigidity of a mature fibril, along with the central amino acid of the terminal molecules, using a force constant of 10 kcal/mol/Å^2. The system was heated to 900 K for 100 ps generating an ensemble of 10 initial states. We then monitored H-bond transitions around the anchored amino acid (Fig. 1a). Pre-nucleation bonds were assessed using an anti-parallel dimer with harmonic restraints on the central amino acids (Fig. 1b). We ran 10 replicas for each starting state lasting for 100 ns each. H-bond transition rates were obtained as described in (14).

All simulations were performed with OpenMM 7.3.1 (24) using three force fields, CHARM36m (25), AMBER99SB-ILDN (26) and AMBER14SB (27), with the TIP3P (28) water model. These force fields are suggested candidates for amyloid peptide assembly based on the study of Aβ_{16-22} dimer in (29). For all simulations, long-range electrostatic interactions were treated with particle mesh Ewald (PME), with both direct-space PME and Lennard-Jones potentials making use of a 10 Å cutoff; the Lennard-Jones potential was switched to zero at the cutoff over a switch width of 1.5 Å to ensure continuity of potential and forces. PME used a relative error tolerance of 10^4 at the cutoff to automatically select the α smoothing parameter, and the default algorithm in OpenMM was used to select Fourier grid spacing (which selected a grid spacing of 0.8 Å in each dimension). All bonds to hydrogen were constrained to a within a fractional error of 10^{-8} of the bond distances using CCMA (30, 31), and waters were rigidly constrained with SETTLE (32). OpenMM's long-range analytical dispersion correction was used to avoid pressure artifacts from truncation of the Lennard-Jones potential. Simulations were run at 300 K with a Monte Carlo barostat with 1 atm external pressure and Monte Carlo update interval of 25 steps. Langevin dynamics was used with a 2 fs time steps and collision of 1 ps^{-1}.

Coordinates were saved every 5 ps and the rate constants between two states, β-sheet and non-β conformations, were

measured using the method in (14). A transition involves either the formation or breakage of a pair of backbone H-bonds between two residues. The distances between amide hydrogen and carbonyl oxygen pair $\mathrm{d_{H-O}}$ were computed using pytraj (33) and monitored for transitions. A backbone H-bond is considered broken when $\mathrm{d_{H-O}}$ exceeds 3.5 Å and formed when $\mathrm{d_{H-O}}$ is shorter than 2.5 Å. In addition to the 1 Å gap in formation and breakage cutoff distances, a five-frame (25 ps) running average of $\mathrm{d_{H-O}}$ and elimination of fast transition less than 300 ps were used to suppress spurious high frequency fluctuations in the detection of backbone H-bond transitions.

To approximate the free energy of the sidechain interactions with the second β-sheet, we repeat the MD simulation using a bilayer of 8 molecules, each of which has 10 amino acids (the pdb structure is taken from Ref. (34)). The two layers are arranged in steric-zipper orientaion in which the side chains of opposing β-sheets are interdigitated and one sheets is shifted upward by half of a strand relative to the other sheet (Fig. 2c,d). The central amino acids of the top and bottom molecules and the six core molecules were harmonically restrained using a force constant of 10 kcal/mol/Å^2. The system was heated to 900 K for 100 ps generating an ensemble of 10 initial states. We ran 10 replicas for each state for 100 ns each and measured H-bond transitions around the anchored amino acid as in the monolayer case. The average time for unbound and bound states are 5272.9 ps and 25649.5 ps respectively, which yield the free energy of $-1.58\ k_{\mathrm{B}}T$.

Lattice simulation details

To observe the kinetics of nucleation we simulate the formation of β-sheets using a lattice model evolving via the Gillespie algorithm (35). Lattice model monomers have 11 amino acids and form anti-parallel β-sheets, which is more stable than parallel β-sheets for polyglutamine (36). Each peptide unit is modeled as a binary variable with states representing β-sheet and non-β conformations. Each amino acid can form a pair of H-bonds, which are mapped to a single bond in the lattice model (Figs. 2). Peptide units at the periphery of the β-core fluctuate between β-sheet and non-β states at rates measured from the all-atom model. New molecules can add to either end of the β-sheet at a concentration-dependent rate approximated by the Smoluchowski formula for an absorbing sphere

$$k_{\mathrm{add}} = 4\pi\sigma D_{\mathrm{m}} c \tag{1}$$

where c is the protein concentration, $\sigma = 1.75\ \mathrm{nm}$ is the radius of the sphere, approximated by half the length of an extended monomer, and D_{m} is the diffusion coefficient of the monomer, $1.79 \times 10^{-10}\mathrm{m}^2/\mathrm{s}$ (34).

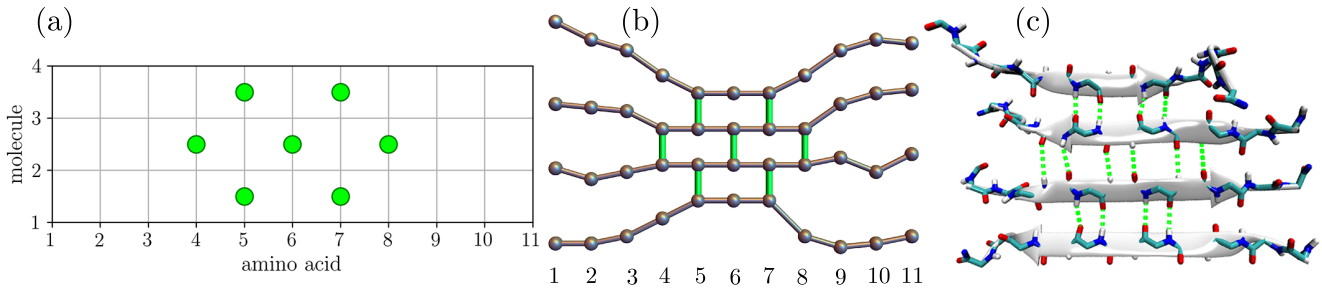

Figure 2: Schematic of the mapping between the lattice and all atom models shown in three representations: (a) lattice (b) ball-and-stick and (c) atomistic (side chains are hidden for clarity). The lattice has a width given by the number of amino acids per molecule and a height given by the current number of molecules in the cluster. Lattice sites (represented by vertical lines) can be occupied, representing an H-bond between connected amino acids, or empty, indicating that at least one of the adjacent amino acids is in the random coil state. The alternating direction of H-bonds (seen in the atomistic view, panel (c)) means that only every-other site can have a bond (green circles in panel (a)).

We define the committor as the probability that a nucleus with a specified number of bonds reaches the cutoff size of 8 molecules before completely dissolving. The committor is calculated as a function of the number of molecules and the total number of H-bonds using 10,000 lattice simulations at each monomer concentration. Each simulation was initiated with a dimer with one H-bond and terminated when there were either 8 molecules in the cluster or all H-bonds were broken. The formation and breakage of H-bonds in the lattice model are described by the hexamer all-atom simulation provided one of the strands remains in the β-state when the bond is broken. Alternatively, if both strands are free to explore non-β states following the breakage of a bond we use the parameters of the dimer simulation. A new molecule can attach to either end of the β-sheet when it has at least two backbone H-bonds.

RESULTS

Simulations show that pre-nucleation bonds are much weaker than post-nucleation bonds.

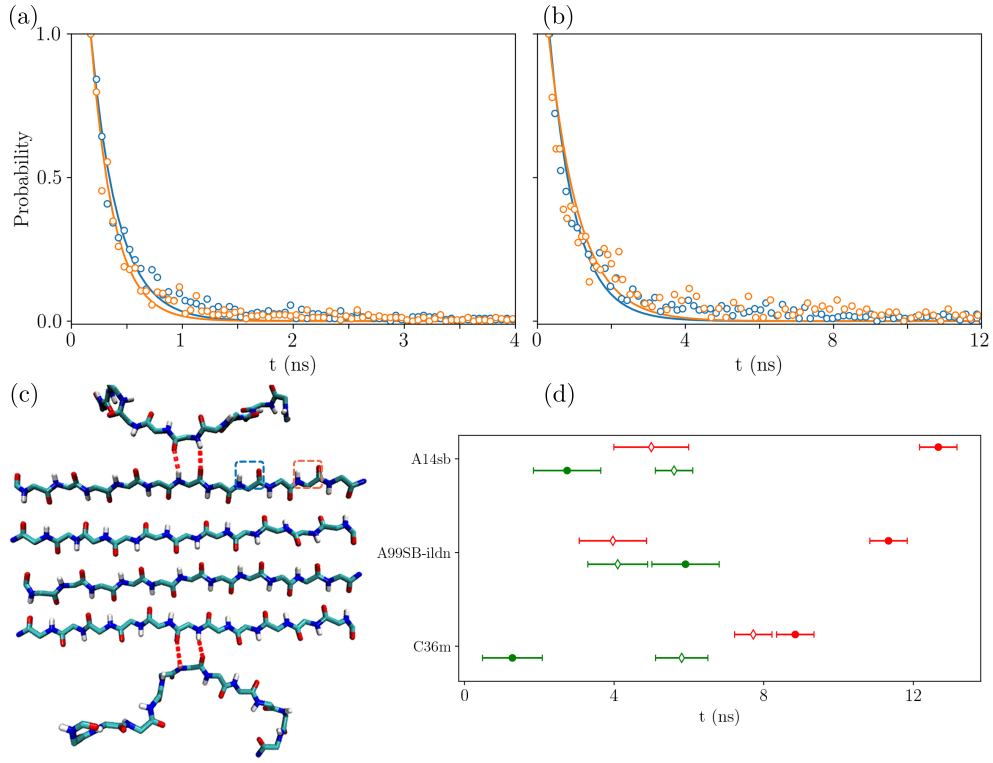

Figure 3: The time distributions of the unbound (a) and bound state (b) of H-bond pairs in the hexamer. Blue points indicate the H-bond pair immediately adjacent to the restrained bonds (red dashes in panel (c)) while orange points indicate the second pair from the restraints, as indicated by the by the highlighted squares in panel (c). Solid lines show exponential fits. (d) Average H-bond pair transition times associated with three different force fields. The open and closed markers represent the times of unbound and bound state of the H-bond pair respectively. Red indicates strong bonds and green indicates weak bonds.

Figure 3 shows the probability that a peptide tail in the all-atom simulation retains the same number of H-bonds as a function of waiting time. These probabilities were fit to an exponential in which the fitting parameter is the bond breakage or formation rate. These rates vary only weakly with position or free chain length allowing us to aggregate the data for better statistics.

We used three force fields to measure the kinetic rate constants. To compare force fields, it is convenient to compute the free energy of the H-bonds using the detailed balance relation

$$\frac{k_{\mathrm{on}}}{k_{\mathrm{off}}} = \mathrm{e}^{-|\varepsilon|/k_{\mathrm{B}}T}. \tag{2}$$

As expected, the β state in the dimer (Fig. 1b), representing the pre-nucleation state, is much less attractive than the post-nucleation hexamer (Fig. 1a). We refer to the bonds in the pre- and post-nucleation states as "weak" and "strong", respectively. Following the expectations of (11) and previous theoretical analysis (37–39) we attribute the difference between these states to the entropic cost of straightening two peptide backbones to form an H-bond in the dimer compared to only one bond that must be immobilized in the post-nucleation case. This explanation is further supported by the fact that the 1.6 k_BT difference between pre- and post-nucleation H-bonds in 2 out of the 3 force fields we tested agrees nearly quantitatively with the $k_{\mathrm{B}}T\ln(6)$ value for the backbone entropy found by fitting helix-coil transitions (40). This additional cost of forming β-sheet structure in the dimer will contribute to the energy barrier that must be surmounted to nucleate a fibril.

Despite the fact that each of the three force fields has been favorably evaluated for simulations of polyglutamine (41–45), they yielded widely different results (Fig. 3d, Table 1). Rather than assess the relative accuracy of these results, we use these different values as representative of sequences in different regimes of fibril stability. AMBER99SB-ILDN is the most attractive,

and in fact, even the weak bonds are net attractive by $\varepsilon_w = -0.37 k_B T$. When the weak bonds are attractive conformational entropy cannot contribute to the barrier. The least attractive force field is CHARMM36m, which was developed in response to overly compact ensembles in simulations of disordered proteins (25). CHARMM36m has a strong bond affinity $\varepsilon_s = -0.14 k_B T$, which has insufficient attraction to observe nucleation within a single β-sheet and suggests that steric zipper interactions would be necessary to form stable fibrils. AMBER14SB has an intermediate affinity with weak bonds $\varepsilon_w = +0.71 k_B T$ that are sufficiently repulsive to create a free energy barrier, yet the strong bonds $\varepsilon_s = -0.93 k_B T$ provide enough attraction to stabilize a single-layered β-sheet.

	C36m	A99SB-ILDN	A14SB
strong bond	-0.14 ± 0.01	-1.05 ± 0.18	-0.93 ± 0.16
weak bond	1.51 ± 0.50	-0.37 ± 0.04	0.71 ± 0.23

Table 1: Strong and weak bond free energies, in units of $k_B T$, calculated from kinetic parameters using Eq. (2). Strong bonds have attractive free energies ranging from -0.14 to -1.05 $k_B T$, while weak bonds range from an attraction of -0.37 $k_B T$ to a repulsion of $+1.51$ $k_B T$.

To gain intuition into how the β-sheet initiation cost affects nucleation kinetics, we first construct a 2D model, consisting of a single β-sheet, for which we can write down the free energy and numerically simulate the kinetics. For this model we employ the parameters from the AMBER14SB force field, which are sufficiently attractive to stabilize an isolated β-sheet, yet repulsive enough to have a barrier.

Model free energy shows that the number of molecules in the critical nucleus is independent of concentration.

We can approximate the free energy of a single β sheet containing $N = \ell(M-1)$ intermolecular H-bonds and M peptides as follows

$$F_1 = N\varepsilon_b + \Delta s(N - M + \ell) + M\mu \tag{3}$$

where ℓ is the average length of a β-strand, ε_b is the favorable contact energy per amino acid including the H-bond and side chain interactions, Δs is the conformational entropy cost to straighten a peptide unit into β conformation, and $\mu = k_B T \ln c/c_0$ is the translational entropy cost of recruiting a molecule from solution, where c is the peptide concentration and c_0 is a reference concentration. The conformational term contains corrections due to the fact that the first H-bond on each molecule does not limit conformational degrees of freedom and the first row of ℓ bonds comes at a cost of $2\Delta s$ per bond. This free energy can be re-written as

$$F_1 = N\varepsilon_s + \ell(\varepsilon_w - \varepsilon_s) + M\mu' \tag{4}$$

where $\varepsilon_s = \varepsilon_b + \Delta s$ is the free energy per H-bond in the interior of the β-sheet, $\varepsilon_w = \varepsilon_b + 2\Delta s$ is the free energy of the H-bonds in a dimer, and $\mu' = \mu - \Delta s$. Note that this analysis neglects minor corrections due to the staggered arrangement of H-bonds, which will be accounted for in the numerical simulation (Fig. 2). If ε_w is negative (attractive), Eq. 4 is monotonically downhill for increasing ℓ. In this regime we recover the well-known result that one-dimensional objects lack a nucleation barrier. Therefore, we focus our attention on repulsive ε_w.

The form of Eq. 4 is that of a bulk term proportional to the sheet area, N, competing against lines tensions that take the values $(\varepsilon_w - \varepsilon_s)/2 = \Delta s/2$ and $\mu'/2$ in the ℓ and M directions, respectively. Notably, the latter line tension is sensitive to the protein concentration. Therefore, we expect that changing the concentration will alter the shape of a nascent β-sheet in the same way that the surface energies of different crystal faces can result in either elongated or compact crystals.

For fixed N, the optimal cluster dimensions are $\ell = (\mu' N/\Delta s)^{1/2}$ and $M = (\Delta s N/\mu')^{1/2} + 1$. These values can be used to find the free energy maximum

$$F_1^{\ddagger} = \mu'\left(\frac{\Delta s}{|\varepsilon_s|} + 1'\right) \tag{5}$$

$$N^{\ddagger} = \frac{\Delta s \mu}{\varepsilon_s^2} \tag{6}$$

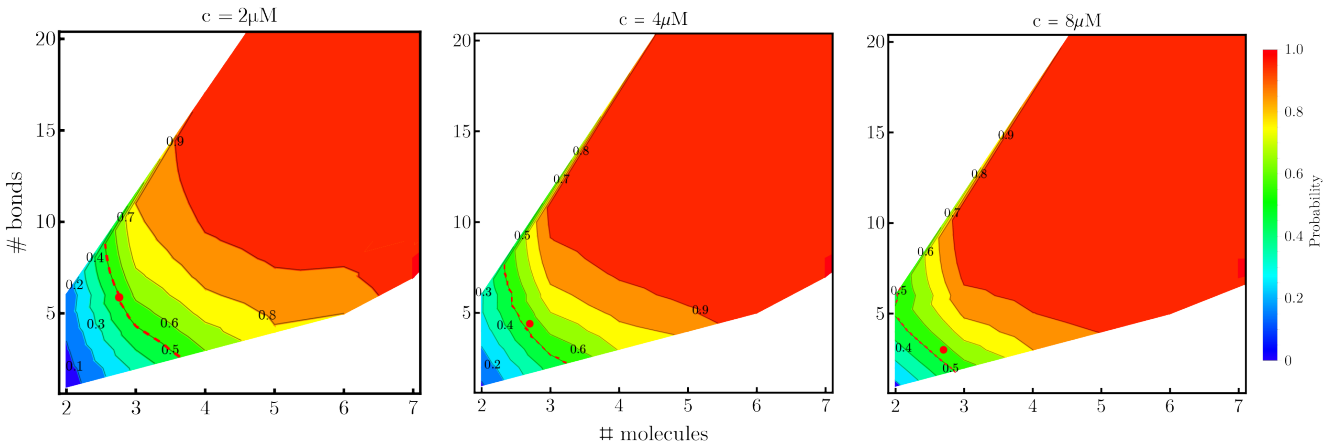

Figure 4: Probability of successful nucleation trajectories with parameters from the AMBER14SB force field as a function of the number of molecules and the total number of H-bonds in the cluster. Increasing concentration shifts the transition surface (50%, red dotted line) to smaller clusters and reduces the need for β-structure due to increased monomer deposition rate. Red dots indicate the free energy saddle points calculated from eqs. 7 and 8.

at which point the cluster has dimensions

$$M^{\ddagger} = \frac{\Delta s}{|\varepsilon_s|} + 1, \tag{7}$$

$$\ell^{\ddagger} = \frac{\mu}{|\varepsilon_s|}. \tag{8}$$

Eq. 7 has a simple interpretation if we notice that the first term is the number of internal bonds needed to pay conformational entropy cost of forming the first weak bond. Both $M^{\ddagger}$ and $\ell^{\ddagger}$ decrease with increasing ε_s, which shows that more aggregation-prone sequences have smaller critical nucleus sizes.

Lattice model simulations show that higher concentrations result in narrower critical nuclei.

Fig. 4 compares the committor, defined as the probability that a nucleus with a specified number of bonds reaches the cutoff size of 8 molecules before completely dissolving, from the kinetic model to the transition state computed from Eqs. 6 and 7. Within error the saddle point lies on the 50% committor line. Hereafter we refer to this line as the critical surface. Consistent with CNT, the critical surface shifts to smaller N as the concentration increases. This can also be seen in Fig. 5a, which plots the average number of H-bonds in successful trajectories crossing the critical surface. The reduction in size is not isotropic and, instead, occurs primarily in the ℓ direction, resulting in shorter β-strands as the concentration increases (Fig. 5b). This can be seen in the inset to Fig. 5b which shows that $N^{\ddagger}/\ell^{\ddagger}$ is nearly constant, consistent with the prediction of Eq. 7 that $M^{\ddagger}$ is independent of concentration. Using the bond free energy from AMBER14SB gives $M^{\ddagger} = 2.76$, which is the number of β strands required to overcome the initial conformational entropy cost of the dimer. From Fig. 5a, we obtain the reference concentration $c_0 = 44.34$ μM by fitting the data to $N^{\ddagger}$ in Eq. 6. This value can be used in the expression for $\ell^{\ddagger} = k_{\mathrm{B}}T\ln(c/c_0)/(\varepsilon_s)$, which matches the simulation results without additional parameters (Fig. 5b).

While $M^{\ddagger}$ is independent of concentration, it plays a crucial role in the state of the nucleus. For $M < M^{\ddagger}$ lateral growth of the β core is unfavorable, which can be seen from the derivative $(\partial F/\partial N)_M > 0$. Therefore, the β-strands tend to remain short, which limits the ability of the nucleus to bind incoming molecules. However, for $M > M^{\ddagger}$, increasing N is favorable, so ℓ will grow by converting the disordered peptide tails to β conformation. The elongation of β-strands increases the size of the template that incoming molecules can bind to. This will increase the binding lifetime of these new molecules, which enables the transition from the nucleation stage to the elongation phase.

To understand the concentration-dependent behavior of the committor, it is instructive to examine the free energy landscape. Expressing Eq. 2 in terms of N and M, we have

$$F_1 = \varepsilon_s N + (\varepsilon_w - \varepsilon_s)N/(M-1) - M\ln(c/c_0) \tag{9}$$

The free energy contours from this expression are plotted in Fig. 6 along with the lowest free energy path. Increasing

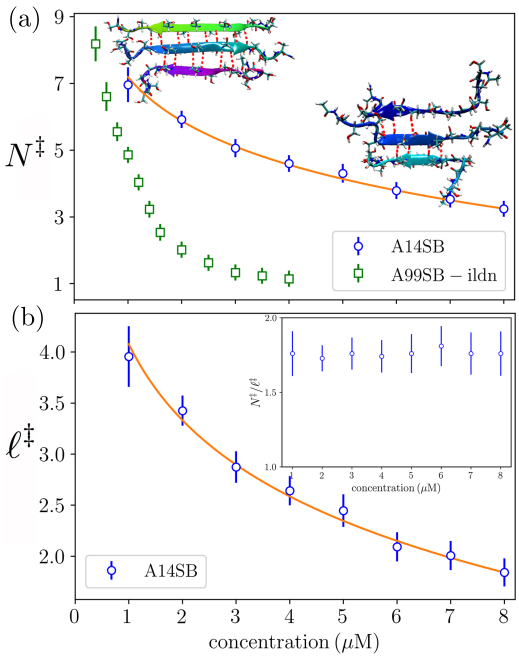

Figure 5: (a) Solution concentration determines the shape of the β-sheet core in critical nuclei as seen by the number of H-bonds in transition clusters (defined by the 50% committor). Line shows fit to Eq. 6. (inset left) Low concentration nuclei have extensive β-structure while higher concentration (inset right) nuclei have shorter β-strands. (b) The number of H-bonds per β-strand in the transition cluster decreases with concentration. Line shows Eq. 8. (inset) The ratio $N^{\ddagger}/\ell^{\ddagger}$ is nearly independent of concentration, consistent with the prediction of Eq. 7 that the number of molecules in the critical cluster does not vary with concentration.

concentration has the effect of shifting the lowest free energy pathway to smaller N, but does not change the number of molecules at the free energy saddle point.

Multi-layered fibrils have a bigger conformational entropy penalty.

The above analysis considers only a single β-sheet, whereas amyloid fibrils contain two or more layered sheets. To study multi-layered fibrils we use our results from the CHARMM36m force field as representative of a peptide that requires two β-sheets for stability. This can be seen from the single sheet parameters $\varepsilon_s = -0.14 k_{\mathrm{B}}T$ and $\varepsilon_w = 1.51 k_{\mathrm{B}}T$, which would require a single sheet length $M \simeq 12$ to overcome the initiation barrier (Eq. 7). Such a large number of molecules incurs a prohibitive translational entropy penalty unless the concentration is extremely high. Therefore, additional interactions, provided by the sidechains, are required to stabilize the fibril.

To assess the effect of the second β-sheet we repeat the MD simulation using a bilayer core of 8 molecules (Fig. 1c,d). Importantly, the staggered interdigitation of sidechains, where the sidechains of molecule i in one sheet insert between the sidechains of molecules i and $i+1$ of the opposite sheet (Fig. 1), implies that molecule additions for $M > 3$ are equivalent. This uniformity allows us to apply the 2D model provided $M \geq 3$.

From the bilayer MD simulations we find that H-bond formation events have a free energy of $-1.58\ k_B T$, which suggests the following breakdown: $\Delta s = \varepsilon_w - \varepsilon_s = 1.65 k_{\mathrm{B}}T$, $\varepsilon_b = \varepsilon_s - \Delta s = -1.79\ k_B T$, and $\varepsilon_{sc} = -1.58 k_{\mathrm{B}}T - \varepsilon_s = -1.44 k_{\mathrm{B}}T$ for the side chain (inter-sheet) interactions. This means that a trimer with two strands in one β-sheet is unfavorable by $(3\Delta s + 2\varepsilon_{sc} + \varepsilon_b)\ell = 0.28 k_{\mathrm{B}}T\ell$, whereas the addition of a fourth molecule results in a state that is favorable by $(4\Delta s + 3\varepsilon_{sc} + 2\varepsilon_b)\ell = -1.3 k_{\mathrm{B}}T\ell$. In this case initiating a new β-sheet with the third molecule is somewhat less favorable than a single sheet trimer, but this configuration allows for a much more favorable tetramer. Generalizing Eq. 3 we can write the free energy of the bilayer as

$$F_2 = (M-2)\ell\varepsilon_b + M(\ell-1)\Delta s + \ell(M-1)\varepsilon_{sc} + M\mu \tag{10}$$

where $N = (M-2)\ell$ is the number of intermolecular H-bonds. After rearranging, Eq. 10 takes a form similar to Eq. 4

$$F_2 = N(\varepsilon_b + \Delta s + \varepsilon_{sc}) + \ell(2\Delta s + \varepsilon_{sc}) + M\mu' \tag{11}$$

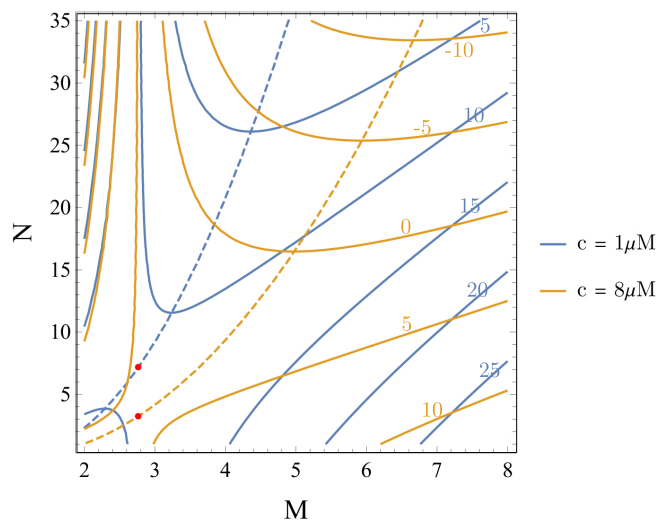

Figure 6: Contour plot of cluster free energy, Eq. 9 as a function of the number of molecules, M, and the number of intermolecular H-bonds, N. Increasing the concentration lowers the overall free energy, decreases the slope of the lowest free energy path (dotted lines), and shifts the free energy saddle point (red dots) to lower values of N.

where the line tension in the ℓ direction has the value $\gamma_2/2 = (2\Delta s + \varepsilon_{sc})/2$. Therefore, a pair of layered β-sheets can also be described using a 2D model in which the β-sheet free energy has line tensions arising from conformational and translational entropy. This conclusion is valid provided the free energy maximum occurs after the initiation of the second β-sheet.

Both Eqs. 4 and 11 have transition state energies given by $M^{\ddagger}\mu'$ implying that the dominant nucleation pathway is simply the one that requires the fewest number of molecules to reach the saddle point. This is given by Eq. 7 for the single sheet pathway and

$$M_2^{\ddagger} = \frac{2\Delta s + \varepsilon_{sc}}{|\varepsilon_b + \Delta s + \varepsilon_{sc}|} + 2 \tag{12}$$

for the two-sheet pathway. Comparison of Eqs. 7 and 12 shows that the two-sheet pathway comes with a greater initial cost because there are two terminal molecules that must be straightened. However, the sidechain interaction energy somewhat reduces the initial cost below $2\Delta s$ and increases the attractive energy of subsequent molecules. The nucleation barrier can be further lowered along the two-sheet pathway if the molecules are long enough to form two β-strands (46).

While our simulations only considered polyglutamine homopolymers, insight into the effect of amino acid sequence can be obtained from the barrier heights given by Eqs. 7 and 12. These expressions show that sidechains that stack favorably, which gives more attractive values of ε_b and ε_{sc}, will reduce the number of β-strands needed to overcome the entropy barrier. To a lesser extent, sidechains can also modulate Δs by perturbing the relative weights of the β and non-β regions of Ramachandran space (47).

DISCUSSION

Experiments have shown that Aβ_{16-22} reduces the lag time of Aβ_{1-40} despite the fact that Aβ_{16-22} molecules are not incorporated in the resulting Aβ_{1-40} fibrils (48). This result is explained by our model, which shows that the β-sheet core contains only a portion of the aggregating molecules. This means that the nucleus will not be sensitive to the molecule length. Therefore, we predict that the lag time reduction by protein mixtures will depend on the protein concentration because at lower concentration the ordered portion of the nucleus will be larger and more sensitive to mismatches. We propose that impurity molecules incorporated during nucleation will be sites susceptible to fragmentation, enabling removal of the defects during elongation.

Our coarse-grained representation only accounts for intermolecular backbone H-bonds, however, we can speculate on the effect of molecules associating with the cluster via non-β contacts. Disordered binding will, in general, lower the free energy of pre-nucleation clusters thereby increasing their prevalence (49). If a β-sheet is to nucleate from disordered clusters (11, 12) there will be an entropy cost to orient molecules with the existing growing β-sheet. This rotational entropy cost will play a similar role as μ in Eq. 3. However, since rotational entropy does not scale with concentration, the aspect ratio of the critical β-sheet will be fixed if the dominant nucleation pathway goes through disordered clusters.

The large role of conformational entropy provides insights into other mechanisms for fibril nucleation, including homogenous nucleation, secondary nucleation catalyzed by fibrils, and heterogeneous nucleation at impurities or interfaces (50–54). According to CNT, binding to a surface during heterogeneous nucleation lowers the surface energy relative to homogeneous nucleation. We propose that the attachment of a peptide to a surface will lower the conformational entropy cost to form a β-strand. Additional evidence for this view comes from the Ramachandran plot, which is dominated by deep wells in the α-helix and β-sheet regions. Therefore, if association with a surface is favorable, the peptide will be strongly biased toward conformations compatible with β-sheet formation. Thus, the association of a peptide with the surface of a fibril will naturally lower the free energy of nucleation.

The nucleation mechanism in our simulations blurs the line between one-step nucleation, where condensation and ordering coincide, and two-step nucleation, where condensation precedes ordering (2, 55–57). We find β-sheet ordering occurs concurrently with initial cluster formation and, in fact, is necessary for molecule retention. However, the molecules remain mostly disordered. Previous simulations of fibril nucleation have shown a two-step mechanism for most cases (11). However, the model in that work represented molecules using an all-or-nothing conversion between ordered and disordered states, hence high concentration or non-specific attractions were necessary to surmount the conversion barrier. Our work shows that a partial conversion to the β-state is an important mechanism to lower the nucleation barrier, so nuclei are likely to have both ordered and disordered regions at all stages.

In conclusion, we have shown that the nucleation of β-sheets, an essential step in the formation of amyloid fibrils, is limited by both conformational and translational free entropy. These factors can be described by classical nucleation theory provided the β-sheet is viewed as a cluster that can grow either by the addition of molecules along the fibril axis or by the conversion of peptide units to the β-state in the lateral direction. We hope that this framework can be used to gain intuition about fibril nucleation in a variety of contexts.

Declaration of Interests The authors declare no competing interests.

Author Contributions T.M.P. and J.D.S. designed the study and wrote the paper. T.M.P. performed the research.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grants R01GM107487 and R01GM141235. J.D.S. acknowledges valuable discussions with Randal Halfmann and Martin Muschol. T.M.P. would like to thank Zhiguang Jia for fruitful discussions. The simulations for this project were performed on the Beocat Research Cluster at Kansas State University, which is funded in part by NSF Grants CHE-1726332, CNS-1006860, EPS-1006860, and EPS-0919443.

REFERENCES

1. Chiti, F., and C. M. Dobson, 2006. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75:333–366.

2. Erdemir, D., A. Y. Lee, and A. S. Myerson, 2009. Nucleation of crystals from solution: classical and two-step models. *Accounts of chemical research* 42:621–629.

3. Chen, J., E. Zhu, J. Liu, S. Zhang, Z. Lin, X. Duan, H. Heinz, Y. Huang, and J. J. De Yoreo, 2018. Building two-dimensional materials one row at a time: Avoiding the nucleation barrier. *Science* 362:1135–1139.

4. Auer, S., C. M. Dobson, M. Vendruscolo, and A. Maritan, 2008. Self-templated nucleation in peptide and protein aggregation. *Physical review letters* 101:258101.

5. Zhang, J., and M. Muthukumar, 2009. Simulations of nucleation and elongation of amyloid fibrils. *The Journal of chemical physics* 130:01B610.

6. Luiken, J. A., and P. G. Bolhuis, 2015. Prediction of a stable associated liquid of short amyloidogenic peptides. *Physical Chemistry Chemical Physics* 17:10556–10567.

7. Luiken, J. A., and P. G. Bolhuis, 2015. Primary nucleation kinetics of short fibril-forming amyloidogenic peptides. *The Journal of Physical Chemistry B* 119:12568–12579.

8. Smit, F. X., J. A. Luiken, and P. G. Bolhuis, 2017. Primary fibril nucleation of aggregation prone Tau fragments PHF6 and PHF6. *The Journal of Physical Chemistry B* 121:3250–3261.

9. Haaga, J., J. Gunton, C. N. Buckles, and J. Rickman, 2018. Early stage aggregation of a coarse-grained model of polyglutamine. *The Journal of chemical physics* 148:045106.

10. Schmit, J. D., K. Ghosh, and K. Dill, 2011. What drives amyloid molecules to assemble into oligomers and fibrils? *Biophysical journal* 100:450–458.

11. Šarić, A., Y. C. Chebaro, T. P. Knowles, and D. Frenkel, 2014. Crucial role of nonspecific interactions in amyloid nucleation. *Proceedings of the National Academy of Sciences* 111:17869–17874.

12. Šarić, A., A. K. Buell, G. Meisl, T. C. Michaels, C. M. Dobson, S. Linse, T. P. Knowles, and D. Frenkel, 2016. Physical determinants of the self-replication of protein fibrils. *Nature physics* 12:874–880.

13. Krausser, J., T. P. Knowles, and A. Šarić, 2020. Physical mechanisms of amyloid nucleation on fluid membranes. *Proceedings of the National Academy of Sciences* 117:33090–33098.

14. Jia, Z., A. Beugelsdijk, J. Chen, and J. D. Schmit, 2017. The Levinthal Problem in Amyloid Aggregation: Sampling of a Flat Reaction Space. *The Journal of Physical Chemistry B* 121:1576–1586.

15. Jia, Z., J. D. Schmit, and J. Chen, 2020. Amyloid assembly is dominated by misregistered kinetic traps on an unbiased energy landscape. *Proceedings of the National Academy of Sciences* 117:10322–10328.

16. Bates, G., P. S. Harper, and L. Jones, 2002. Huntington's disease. 45. Oxford University Press, USA.

17. Arrasate, M., and S. Finkbeiner, 2012. Protein aggregates in Huntington's disease. *Experimental neurology* 238:1–11.

18. Zoghbi, H. Y., and H. T. Orr, 2000. Glutamine repeats and neurodegeneration. *Annual review of neuroscience* 23:217–247.

19. Walker, F. O., 2007. Huntington's disease. *The Lancet* 369:218–228.

20. Kar, K., M. Jayaraman, B. Sahoo, R. Kodali, and R. Wetzel, 2011. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nature structural & molecular biology* 18:328.

21. Chen, S., F. A. Ferrone, and R. Wetzel, 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proceedings of the National Academy of sciences* 99:11884–11889.

22. Chen, M., M. Tsai, W. Zheng, and P. G. Wolynes, 2016. The aggregation free energy landscapes of polyglutamine repeats. *Journal of the American Chemical Society* 138:15197–15203.

23. Chen, M., and P. G. Wolynes, 2017. Aggregation landscapes of Huntingtin exon 1 protein fragments and the critical repeat length for the onset of Huntington's disease. *Proceedings of the National Academy of Sciences* 114:4406–4411.

24. Eastman, P., J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, et al., 2017. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* 13:e1005659.

25. Huang, J., S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* 14:71–73.

26. Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* 78:1950–1958.

27. Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* 11:3696–3713.

28. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* 79:926–935.

29. Man, V. H., X. He, P. Derreumaux, B. Ji, X.-Q. Xie, P. H. Nguyen, and J. Wang, 2019. Effects of all-atom molecular mechanics force fields on amyloid peptide assembly: the case of aβ16–22 dimer. *Journal of chemical theory and computation* 15:1440–1452.

30. Barth, E., K. Kuczera, B. Leimkuhler, and R. D. Skeel, 1995. Algorithms for constrained molecular dynamics. *Journal of computational chemistry* 16:1192–1209.

31. Eastman, P., and V. S. Pande, 2010. Constant constraint matrix approximation: a robust, parallelizable constraint method for molecular simulations. *Journal of chemical theory and computation* 6:434–437.

32. Miyamoto, S., and P. A. Kollman, 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* 13:952–962.

33. Roe, D. R., and T. E. Cheatham III, 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* 9:3084–3095.

34. Punihaole, D., R. S. Jakubek, R. J. Workman, L. E. Marbella, P. Campbell, J. D. Madura, and S. A. Asher, 2017. Monomeric polyglutamine structures that evolve into fibrils. *The Journal of Physical Chemistry B* 121:5953–5967.

35. Gillespie, D. T., 1977. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry* 81:2340–2361.

36. Punihaole, D., R. J. Workman, Z. Hong, J. D. Madura, and S. A. Asher, 2016. Polyglutamine fibrils: new insights into antiparallel β-sheet conformational preference and side chain structure. *The Journal of Physical Chemistry B* 120:3012–3026.

37. Zhang, L., and J. D. Schmit, 2016. Pseudo-one-dimensional nucleation in dilute polymer solutions. *Physical Review E* 93:060401.

38. Zhang, L., and J. D. Schmit, 2017. Theory of amyloid fibril nucleation from folded proteins. *Israel journal of chemistry* 57:738–749.

39. Lee, C.-T., and E. M. Terentjev, 2017. Mechanisms and rates of nucleation of amyloid fibrils. *The Journal of chemical physics* 147:105103.

40. Ghosh, K., and K. Dill, 2009. Theory for protein folding cooperativity: helix bundles. *Journal of the American Chemical Society* 131:2306–2312.

41. Fluitt, A. M., and J. J. de Pablo, 2015. An analysis of biomolecular force fields for simulations of polyglutamine in solution. *Biophysical journal* 109:1009–1018.

42. Wen, J., D. R. Scoles, and J. C. Facelli, 2017. Molecular dynamics analysis of the aggregation propensity of polyglutamine segments. *PloS one* 12.

43. Gómez-Sicilia, À., M. Sikora, M. Cieplak, and M. Carrión-Vázquez, 2015. An exploration of the universe of polyglutamine structures. *PLoS computational biology* 11.

44. Cote, S., G. Wei, and N. Mousseau, 2012. All-atom stability and oligomerization simulations of polyglutamine nanotubes with and without the 17-amino-acid N-terminal fragment of the Huntingtin protein. *The journal of physical chemistry B* 116:12168–12179.

45. Starikov, E., H. Lehrach, and E. Wanker, 1999. Folding of oligoglutamines: a theoretical approach based upon thermodynamics and molecular mechanics. *Journal of Biomolecular Structure and Dynamics* 17:409–427.

46. Kandola, T., J. Zhang, S. Venkatesan, B. Lerbakken, J. F. Blanck, J. Wu, J. Unruh, P. Berry, J. L. Lange, A. Von Schulze, et al., 2021. The polyglutamine amyloid nucleus in living cells is monomeric and has competing dimensions of order. *bioRxiv* .

47. Chakrabarti, P., and D. Pal, 2001. The interrelationships of side-chain and main-chain conformations in proteins. *Progress in biophysics and molecular biology* 76:1–102.

48. Bunce, S. J., Y. Wang, K. L. Stewart, A. E. Ashcroft, S. E. Radford, C. K. Hall, and A. J. Wilson, 2019. Molecular insights into the surface-catalyzed secondary nucleation of amyloid-β40 (Aβ40) by the peptide fragment Aβ16–22. *Science Advances* 5:eaav8216.

49. Phan, T. M., S. Whitelam, and J. D. Schmit, 2019. Catalystlike role of impurities in speeding layer-by-layer growth. *Physical Review E* 100:042114.

50. Törnquist, M., T. C. Michaels, K. Sanagavarapu, X. Yang, G. Meisl, S. I. Cohen, T. P. Knowles, and S. Linse, 2018. Secondary nucleation in amyloid formation. *Chemical Communications* 54:8667–8684.

51. Xue, W.-F., S. W. Homans, and S. E. Radford, 2008. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proceedings of the National Academy of Sciences* 105:8926–8931.

52. Cohen, S. I., S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson, and T. P. Knowles, 2013. Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proceedings of the National Academy of Sciences* 110:9758–9763.

53. Meisl, G., X. Yang, E. Hellstrand, B. Frohm, J. B. Kirkegaard, S. I. Cohen, C. M. Dobson, S. Linse, and T. P. Knowles, 2014. Differences in nucleation behavior underlie the contrasting aggregation kinetics of the Aβ40 and Aβ42 peptides. *Proceedings of the National Academy of Sciences* 111:9384–9389.

54. Cohen, S. I., R. Cukalevski, T. C. Michaels, A. Šarić, M. Törnquist, M. Vendruscolo, C. M. Dobson, A. K. Buell, T. P. Knowles, and S. Linse, 2018. Distinct thermodynamic signatures of oligomer generation in the aggregation of the amyloid-β peptide. *Nature chemistry* 10:523.

55. ten Wolde, P. R., and D. Frenkel, 1997. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* 277:1975–1978.

56. Vekilov, P. G., 2004. Dense liquid precursor for the nucleation of ordered solid phases from solution. *Crystal Growth & Design* 4:671–685.

57. Filobelo, L. F., O. Galkin, and P. G. Vekilov, 2005. Spinodal for the solution-to-crystal phase transformation. *The Journal of chemical physics* 123:014904.