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

Tien M. Phan, and Jeremy D. Schmit*
Department of Physics, Kansas State University, Manhattan, KS 66506, USA

The formation of amyloid fibrils in Alzheimer's disease and other neurodegenerative disorders is limited by a slow nucleation step due to the entropic cost to initiate the ordered cross-β structure. While the barrier can be lowered if the molecules maintain conformational disorder, poorly ordered clusters provide a poor binding surface for new molecules. To understand these opposing factors, we used all-atom simulations to parameterize a lattice model that treats each amino acid as a binary variable with β-sheet and non-β states. We find that the optimal degree of order in a nucleus depends on protein concentration. Low concentration systems require more ordered nuclei to capture infrequent monomer attachments. The nucleation phase transitions to the elongation phase when the β-sheet core becomes large enough to overcome the initiation cost, at which point further ordering becomes favorable and the nascent fibril efficiently captures new molecules.

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 to the fibril state is limited by nucleation events. Nucleation processes may be homogenous events, which occur in free solution, and heterogeneous events, which are catalyzed by surfaces or other species in the solution. An important class of the latter events is secondary nucleation, in which existing fibrils promote the formation of new ones [2–6]. The combined effect of these nucleation events is an exponential acceleration in the monomer conversion rate.

Nucleation behavior indicates a free energy barrier separating the soluble and condensed states. Heterogeneous and secondary nucleation often dominate over homogeneous nucleation because they lower the barrier [7]. In the case of isotropic particle condensation, the barrier arises because particles at the periphery have sacrificed the translational entropy of the dilute phase, but only have a fraction of the favorable interactions available to interior particles. 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 [8]. Early attempts to explain amyloid nucleation identified β-sheet layering as a second assembly dimension [9, 10]. Another contribution to the barrier arises from the conformational entropy cost of straightening into a β-strand. For a molecule joining an established fibril the template is rigid and entropy loss is limited to the incoming molecule. However, the initiation of β-structure from two disordered molecules requires both molecules to lose conformational entropy. As a result, this event is net repulsive [11, 12]. This means that amyloid nucleation is better viewed as a 3D incorporation of amino acid building blocks rather than a 1D or 2D assembly of polypeptide strands. The β-sheet must be 4-6 strands thick to overcome the entropic cost of straightening the first molecule [12], consistent with numerous simulations showing that fibrils shorter than this size are unstable [13–21].

Previously, we used a three-molecule toy model to investigate conformational conversion during nucleation, [11]. This model revealed that conversion is a compromise between minimizing free energy by maintaining disorder and the retention of incoming molecules with a large β-sheet template.

In this Letter we examine the more realistic scenario where > 3 molecules are needed to surmount the barrier. To access nucleation timescales, we adopt a multi-scale approach in which all-atom simulations are used to parameterize a lattice model. We find that critical nuclei consist of a β-sheet core surrounded by disordered tails. The shape of the ordered core depends on both the aggregation propensity and the protein concentration. Concentrated solutions favor clusters with shorter β-strands, while lower concentrations favor longer β-strands. During the transition from nucleation to elongation β-sheet order propagates along the molecules until the β-strands reach their maximum length. This enlarges the available template, which is responsible for the transition from transient binding during nucleation to nearly irreversible binding during elongation.

Lattice model captures molecule addition and β-sheet formation. Our model is based on a Markov State Model (MSM) developed to study fibril elongation [22, 23]. In those works the conformational search was discretized using two reaction coordinates: 1) the alignment of an incoming molecule with the template and 2) the number of β-sheet H-bonds. Here we remove the alignment complication by considering polyglutamine, motivated by the aggregation-prone region of huntingtin protein [24–32].

The lattice model, illustrated in Fig. 1(a), evolves by the Gillespie algorithm [33]. Lattice model monomers

*schmit@phys.ksu.edu

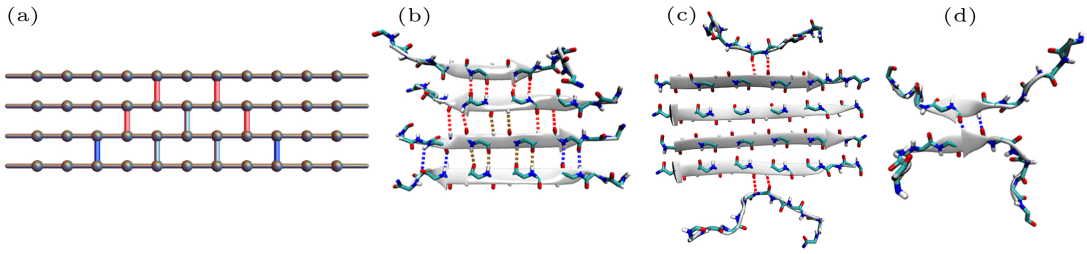

FIG. 1: (a) Snapshot from the lattice model mapped to (b) the corresponding molecular representation. Bonds are color coded by type. Strong = red, weak = blue, and grey bonds are immobilized by other bonds. (c, d) MD snapshots from the sampling of strong and weak H-bonds (side chains are not shown for clarity). Kinetic parameters for strong bonds are sampled from the terminal molecules on an established cluster (c), while weak bonds are sampled using a dimer that is harmonically restrained at the central amino acids (d).

have 11 amino acids and form anti-parallel β-sheets. This is more stable than parallel β-sheets for polyglutamine [34]. 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 [Fig. 1(a, b)]. 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

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

where c is the protein concentration, σ 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, $1.79 \times 10^{-10}\text{m}^2/\text{s}$ [35].

The committor is used to identify the transition state ensemble. The committor is defined as the probability that a nucleus of a specific size reaches the cutoff size of 15 molecules before completely dissolving. If a molecule breaks all bonds with the cluster it is considered to return to free solution. Our model does not consider molecules that associate with the cluster without backbone H-bonds. These non-β contacts are a negligible contribution to the association time during elongation [23]. The lower β-content during nucleation will increase the importance of non-β states, however, the effect amounts to concentration rescaling by helping to recruit molecules. The effect of non-β states increases at high concentrations where disordered oligomers are stable.

H-bond formation and breakage rates are obtained from all-atom simulations. We use two sets of bonding rate constants, depending on the position in the cluster. *"Strong"* bonds form between a disordered amino acid and an established β-strand. *"Weak"* bonds form between two disordered amino acids. These bonds result in the loss of conformational entropy from both backbones, a greater penalty than bonds with a pre-existing strand. Strong and weak bond kinetic parameters were assessed using all-atom simulations on the hexamer and dimer assemblies shown in Fig. 1(c, d) with harmonic restraints applied to the non-sampled bonds (see SI). Here we present results for the AMBER14SB force field, which has intermediate affinity of the three we tested (see SI). AMBER14SB weak bonds have free energy $\epsilon_w = 0.71 k_{\text{B}}T$ and the strong bonds, $\epsilon_s = -0.93 k_{\text{B}}T$, provide enough attraction to stabilize a single-layered β-sheet.

Established β-structure helps capture new molecules. The assembly will grow when the attachment rate is greater than the detachment rate and shrink when the detachment rate is greater. While the attachment rate is a function of concentration (Eq. 1), detachment is limited by the rupture of favorable interactions. To gain intuition, consider an Arrhenius model in which the detachment rate scales as $e^{-n_s|\epsilon_s|/k_{\text{B}}T}$, where n_s is the number of strong bonds between the cluster and the departing molecule. In the dimer $n_s = 0$, so the detachment rate is large. In the elongation phase n_s is given by the molecule length so the detachment rate is small.

Fibril nucleation is slow because it is unfavorable to form a template large enough to capture new molecules. However, the attachment rate increases with concentration, reducing the required template size. This is seen in Fig. 2 which shows the committor as a function of nucleus size. At high concentration (Fig. 2c) nucleation is more likely than dissolution for a cluster containing about four molecules and ten H-bonds. But, at lower concentration (Fig. 2a) a 50% committor is not reached until the cluster has about eight molecules and 30 H-bonds. This trend toward larger nuclei at lower concentrations is consistent with classical nucleation theory (CNT). Note that the slope of the 50% committor changes with concentration. At high concentration the 50% contour is nearly vertical, indicating that the *number of molecules* in the cluster is the primary determinant of nucleation probability. In this case, molecule addition is sufficiently fast that an ordered template is not necessary to retain molecules. Once the cluster reaches $\sim$ 4 molecules, it becomes resistant to the transient loss of monomers and, most likely, will continue to grow.

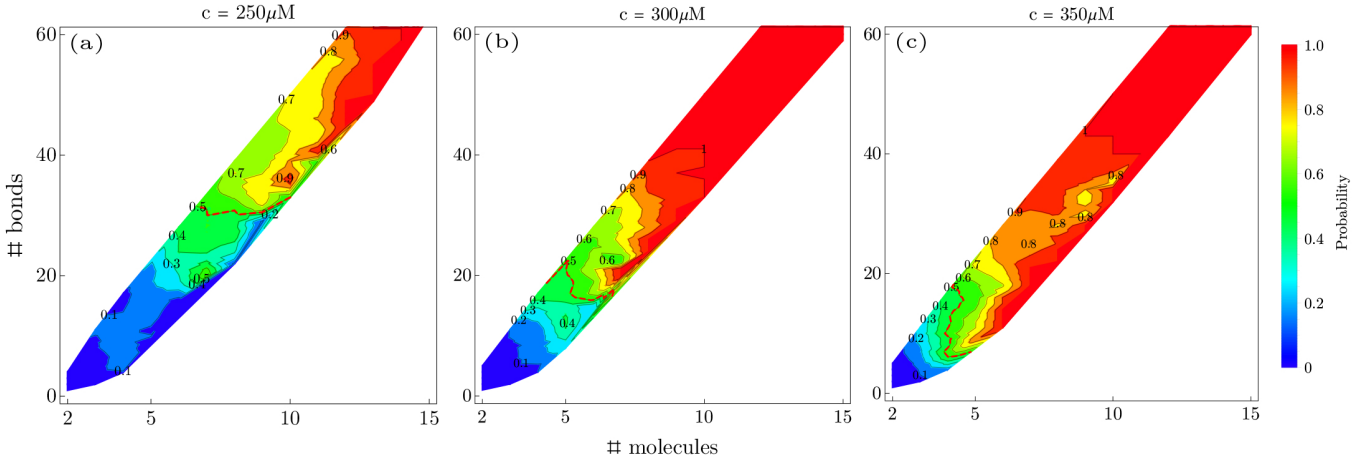

FIG. 2: Probability of successful nucleation trajectories with parameters from the AMBER14SB force field as a function of the number of molecules and the total number of H-bonds in the cluster. Increasing the concentration shifts the 50% probability contour (red dotted line) to smaller clusters and reduces the need for β-structure due to the increased monomer deposition rate.

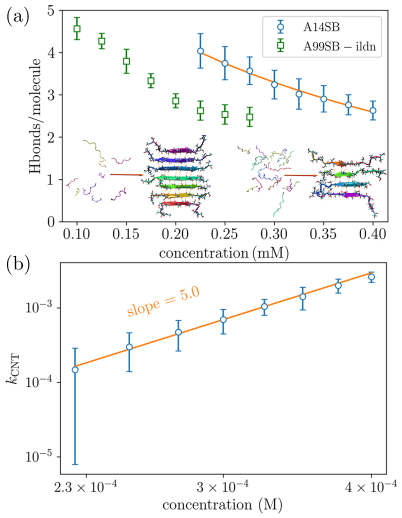

FIG. 3: (a) Solution concentration determines the shape of the β-sheet core in critical nuclei as seen by the number of H-bonds per molecule in the critical clusters (defined by the 50% committor). Line shows fit to Eq. 5. (inset left) Low concentration nuclei have extensive β-structure to provide a strong binding surface for newly docked molecules. (inset right) Higher concentration nuclei have shorter β-strands because the higher deposition rate places lower demands on retaining new additions. (b) Nucleation success rate versus monomer concentration measured in lattice simulations.

In contrast, the 50% committor is nearly horizontal at lower concentration (Fig. 2a), indicating that the *degree of structural order* is more important than the number of molecules. This is because monomer addition is infrequent, so successful nuclei are those that retain added molecules, which requires large n_s. Note that structural ordering is not a consequence of longer waiting times. Acquiring order is unfavorable below the critical size, so successful nuclei are the small fraction with fluctuations taking them uphill on the free energy landscape.

Nucleation at low concentration requires highly ordered clusters. The effects of concentration are highlighted in Fig. 3a, which shows the number of H-bonds per molecule in clusters on the 50% committor line. Again, we see the trend that low c requires a high degree of order ($\sim$ 4 bonds per molecule) in critical clusters, while poorly ordered clusters ($\sim$ 2.5 bonds/molecule) are enough at high c. Also, the concentration behavior depends on the aggregation propensity of the molecules. This can be seen from the more aggregation-prone AMBER99SB-ILDN force field, which shows high concentration behavior at 0.2 mM (Fig. 3a). In contrast, AMBER14SB has a larger detachment rate and requires a $c \simeq 0.4$ mM to nucleate from low order clusters.

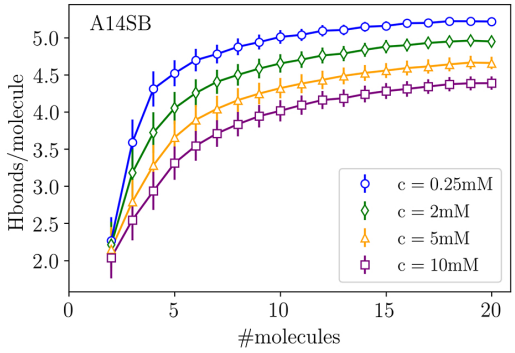

FIG. 4: The free energy cost of forming β-structure can be overcome by the formation of strong bonds in the interior of a β-sheet. As the cluster grows larger, these bonds promote an ordering transition.

These results can be understood within CNT using the

cluster free energy

$$F = N\epsilon_s + \gamma_b b + \gamma_h h, \tag{2}$$

where b and h are the β-sheet dimensions in the β-strand and molecule stacking directions, respectively, the vertical surface tension is given by the translational entropy cost to recruit a molecule $\gamma_h = -k_B T \ln c/c_0$ (c_0 is a reference concentration), the horizontal surface tension is $\gamma_b = \epsilon_w - \epsilon_s$, and the total number of bonds in the cluster is $N = abh$, where a is a geometric factor ($a = 1$ for a rectangular cluster). For fixed N, the optimal cluster dimensions are $b = (\gamma_h N/a\gamma_b)^{1/2}$ and $h = (\gamma_b N/a\gamma_h)^{1/2}$, which can be used to find the free energy maximum

$$F^\ddagger = \frac{\gamma_b\gamma_h}{a|\epsilon_s|}, \quad N^\ddagger = \frac{\gamma_b\gamma_h}{a\epsilon_s^2} \tag{3}$$

at which point the cluster has dimensions

$$h^\ddagger = \frac{\gamma_b}{a|\epsilon_s|}, \tag{4}$$

$$b^\ddagger = \frac{\gamma_h}{a|\epsilon_s|}. \tag{5}$$

The nucleation rate from CNT is

$$k_{\mathrm{CNT}} \sim c\mathrm{e}^{-F^\ddagger/k_B T} = c\mathrm{e}^{-h^\ddagger\gamma_h/k_B T}, \tag{6}$$

where the Arrhenius term gives the probability of finding a cluster at the free-energy maximum $F^\ddagger$. From Eq. 6, we can verify that $h^\ddagger$ satisfies the relationship

$$h^\ddagger = \frac{d\ln(k_{\mathrm{CNT}})}{d\ln c} - 1 = n_c - 1, \tag{7}$$

where n_c is the effective reaction order of nucleation [36]. Eq. 7 shows that the nucleus size can be obtained from the slope of double logarithmic plot of the nucleation rate against monomer concentration [36]. This plot is shown in Fig. 3b, with k_{CNT} given by the fraction of simulations that reach the cutoff size of 15 molecules.

From Fig. 3b we obtain $n_c = h^\ddagger + 1 = 5.0$, which can be used in Eq. 5 to find $a = 0.44$, consistent with a diamond-shaped β-sheet. This value can be used in the expression for $b^\ddagger = k_B T \ln(c/c_0)/(a\epsilon_s)$, which captures the change in β-ordering as a function of concentration (Fig. 3a). Here the reference concentration $c_0 = 1.15$ mM, determined by fitting, sets the vertical offset.

Nucleation transitions to elongation when the cluster is large enough to spontaneously order. The defining feature of the nucleation phase is that it is dominated by the entropic cost of straightening the first molecule [11]. When the cluster reaches 4-6 β-strands strong bonds within the β-sheet provide enough energy to overcome this cost. At this point it is no longer advantageous for the cluster to maintain disordered tails and β-structure will extend along the entirety of the molecules. Fig. 4 shows the concentration dependence of the transition from nucleation to elongation. At high c ordering happens more gradually than low c, as seen by a lower number of H-bonds per molecule. Two factors contribute to this result. First, ordering does not need to be as extensive at high c to achieve the capture efficiency of elongation. Secondly, elongation at high c outpaces the ordering of newly attached molecules.

Disorder in the nucleation phase allows promiscuity in cross-seeding. Aβ_{16-22} reduces the lag time of Aβ_{1-40} despite the fact that molecules are not mixed in the resulting fibrils [37]. This is explained by our results showing that the critical nucleus contains only a portion of the aggregating molecules, so the nucleus will not be sensitive to molecule length. Therefore, the lag time reduction by protein mixtures will depend on c because at lower c the ordered portion of the nucleus will be larger and more sensitive to mismatches.

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 [7, 38–40]. 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 [36]. However, the model in that work represented molecules using an all-or-nothing conversion between ordered and disordered states. Our work shows that a partial conversion to the ordered state is an important mechanism to lower the nucleation barrier.

Our coarse-grained representation only accounts for intermolecular backbone H-bonds, allowing for analysis of secondary structure formation. However, sidechain and non-β backbone contacts are potentially significant in at least two ways. First, omitting non-β interactions prevents completely disordered clusters. This is less of an issue at concentrations below the critical concentration for disordered oligomers. But, disordered binding will, in general, lower the free energy of pre-nucleation clusters [41]. Second, our model does not include the stacking of β-sheets via steric zipper interactions [42]. The absence of single-layer fibrils in experiments suggests that steric zippers are necessary for fibril stability. This implies that the stability of single-layer sheets in our simulations is a result of force fields that overly stabilize protein-protein interactions [43–46]. Should this be a force field artifact, it is fortuitous for our study because it allows us to study β-sheet initiation without the complication of multiple layers.

Conclusion – There are many different mechanisms for fibril nucleation, including homogenous nucleation, secondary nucleation catalyzed by fibrils, and heterogeneous nucleation at impurities or interfaces. All of these pathways must surmount a free energy barrier that arises from the fact that immature clusters lack the stabilizing interactions of established fibrils so the entropic costs of condensing and ordering are incompletely compensated.

Our work shows that the conformational entropy contribution to the barrier is reduced by limiting the extent of secondary structure in the cluster and that the optimal amount of structure depends on the concentration of free protein. This finding should apply regardless of the specific nucleation pathway.

This work was supported by National Institutes of Health Grant R01GM107487. J.D.S. acknowledges valuable discussions with Randal Halfmann. 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.

[1] F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).
[2] M. Törnquist, T. C. Michaels, K. Sanagavarapu, X. Yang, G. Meisl, S. I. Cohen, T. P. Knowles, and S. Linse, Chemical Communications **54**, 8667 (2018).
[3] W.-F. Xue, S. W. Homans, and S. E. Radford, Proceedings of the National Academy of Sciences **105**, 8926 (2008).
[4] S. I. Cohen, S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson, and T. P. Knowles, Proceedings of the National Academy of Sciences **110**, 9758 (2013).
[5] G. Meisl, X. Yang, E. Hellstrand, B. Frohm, J. B. Kirkegaard, S. I. Cohen, C. M. Dobson, S. Linse, and T. P. Knowles, Proceedings of the National Academy of Sciences **111**, 9384 (2014).
[6] S. I. Cohen, R. Cukalevski, T. C. Michaels, A. Šarić, M. Törnquist, M. Vendruscolo, C. M. Dobson, A. K. Buell, T. P. Knowles, and S. Linse, Nature chemistry **10**, 523 (2018).
[7] D. Erdemir, A. Y. Lee, and A. S. Myerson, Accounts of chemical research **42**, 621 (2009).
[8] J. Chen, E. Zhu, J. Liu, S. Zhang, Z. Lin, X. Duan, H. Heinz, Y. Huang, and J. J. De Yoreo, Science **362**, 1135 (2018).
[9] S. Auer, C. M. Dobson, M. Vendruscolo, and A. Maritan, Physical review letters **101**, 258101 (2008).
[10] J. Zhang and M. Muthukumar, The Journal of chemical physics **130**, 01B610 (2009).
[11] L. Zhang and J. D. Schmit, Physical Review E **93**, 060401 (2016).
[12] L. Zhang and J. D. Schmit, Israel journal of chemistry **57**, 738 (2017).
[13] J. Nasica-Labouze and N. Mousseau, PLoS computational biology **8** (2012).
[14] U. F. Röhrig, A. Laio, N. Tantalo, M. Parrinello, and R. Petronzio, Biophysical journal **91**, 3217 (2006).
[15] R. D. Hills Jr and C. L. Brooks III, Journal of molecular biology **368**, 894 (2007).
[16] Y. Zou, Y. Sun, Y. Zhu, B. Ma, R. Nussinov, and Q. Zhang, ACS chemical neuroscience **7**, 286 (2016).
[17] T. T. Tran, P. H. Nguyen, and P. Derreumaux, The Journal of chemical physics **144**, 205103 (2016).
[18] J. Haaga, J. Gunton, C. N. Buckles, and J. Rickman, The Journal of chemical physics **148**, 045106 (2018).
[19] N. T. Co and M. S. Li, The Journal of chemical physics **137**, 095101 (2012).
[20] J. A. Luiken and P. G. Bolhuis, Physical Chemistry Chemical Physics **17**, 10556 (2015).
[21] F. X. Smit, J. A. Luiken, and P. G. Bolhuis, The Journal of Physical Chemistry B **121**, 3250 (2017).
[22] Z. Jia, A. Beugelsdijk, J. Chen, and J. D. Schmit, The Journal of Physical Chemistry B **121**, 1576 (2017).
[23] Z. Jia, J. D. Schmit, and J. Chen, Proceedings of the National Academy of Sciences **117**, 10322 (2020).
[24] G. Bates, P. S. Harper, and L. Jones, *Huntington's disease*, 45 (Oxford University Press, USA, 2002).
[25] M. Arrasate and S. Finkbeiner, Experimental neurology **238**, 1 (2012).
[26] H. Y. Zoghbi and H. T. Orr, Annual review of neuroscience **23**, 217 (2000).
[27] F. O. Walker, The Lancet **369**, 218 (2007).
[28] K. Kar, M. Jayaraman, B. Sahoo, R. Kodali, and R. Wetzel, Nature structural & molecular biology **18**, 328 (2011).
[29] S. Chen, F. A. Ferrone, and R. Wetzel, Proceedings of the National Academy of sciences **99**, 11884 (2002).
[30] M. Chen, M. Tsai, W. Zheng, and P. G. Wolynes, Journal of the American Chemical Society **138**, 15197 (2016).
[31] M. Chen and P. G. Wolynes, Proceedings of the National Academy of Sciences **114**, 4406 (2017).
[32] T. T. Phan and J. D. Schmit, Biophysical Journal (2020).
[33] D. T. Gillespie, The journal of physical chemistry **81**, 2340 (1977).
[34] D. Punihaole, R. J. Workman, Z. Hong, J. D. Madura, and S. A. Asher, The Journal of Physical Chemistry B **120**, 3012 (2016).
[35] D. Punihaole, R. S. Jakubek, R. J. Workman, L. E. Marbella, P. Campbell, J. D. Madura, and S. A. Asher, The Journal of Physical Chemistry B **121**, 5953 (2017).
[36] A. Šarić, A. K. Buell, G. Meisl, T. C. Michaels, C. M. Dobson, S. Linse, T. P. Knowles, and D. Frenkel, Nature physics **12**, 874 (2016).
[37] S. J. Bunce, Y. Wang, K. L. Stewart, A. E. Ashcroft, S. E. Radford, C. K. Hall, and A. J. Wilson, Science Advances **5**, eaav8216 (2019).
[38] P. R. ten Wolde and D. Frenkel, Science **277**, 1975 (1997).
[39] P. G. Vekilov, Crystal Growth & Design **4**, 671 (2004).
[40] L. F. Filobelo, O. Galkin, and P. G. Vekilov, The Journal of chemical physics **123**, 014904 (2005).
[41] T. M. Phan, S. Whitelam, and J. D. Schmit, Physical Review E **100**, 042114 (2019).
[42] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, et al., Nature **447**, 453 (2007).
[43] M. R. Jensen and M. Blackledge, Proceedings of the National Academy of Sciences **111**, E1557 (2014).
[44] R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. MacKerell Jr, Journal of chemical theory and computation **8**, 3257 (2012).
[45] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, The journal of physical chemistry B **119**, 5113 (2015).
[46] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, Nature methods **14**, 71 (2017).