Amyloid assembly is dominated by misregistered kinetic traps on an unbiased energy landscape

Zhiguang Jia[a], Jeremy D. Schmit[b,1], and Jianhan Chen[a,c,1]

[a]Department of Chemistry, University of Massachusetts, Amherst, MA 01003; [b]Department of Physics, Kansas State University, Manhattan, KS 66506; and [c]Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, MA 01003

Edited by Rohit V. Pappu, Washington University in St. Louis, St. Louis, MO, and accepted by Editorial Board Member Susan Marqusee March 13, 2020 (received for review June 28, 2019)

Atomistic description of protein fibril formation has been elusive due to the complexity and long time scales of the conformational search. Here, we develop a multiscale approach combining numerous atomistic simulations in explicit solvent to construct Markov State Models (MSMs) of fibril growth. The search for the in-register fully bound fibril state is modeled as a random walk on a rugged two-dimensional energy landscape defined by β-sheet alignment and hydrogen-bonding states, whereas transitions involving states without hydrogen bonds are derived from kinetic clustering. The reversible association/dissociation of an incoming peptide and overall growth kinetics are then computed from MSM simulations. This approach is applied to derive a parameter-free, comprehensive description of fibril elongation of $A\beta_{16-22}$ and how it is modulated by phenylalanine-to-cyclohexylalanine (CHA) mutations. The trajectories show an aggregation mechanism in which the peptide spends most of its time trapped in misregistered β-sheet states connected by weakly bound states twith short lifetimes. Our results recapitulate the experimental observation that mutants CHA19 and CHA1920 accelerate fibril elongation but have a relatively minor effect on the critical concentration for fibril growth. Importantly, the kinetic consequences of mutations arise from cumulative effects of perturbing the network of productive and nonproductive pathways of fibril growth. This is consistent with the expectation that nonfunctional states will not have evolved efficient folding pathways and, therefore, will require a random search of configuration space. This study highlights the importance of describing the complete energy landscape when studying the elongation mechanism and kinetics of protein fibrils.

amyloid | Markov State Model | aggregation | molecular dynamics

Aggregation of misfolded proteins has been implicated in many devastating and currently incurable medical disorders such as Alzheimer's and prion diseases (1–3). In these diseases, the misfolded protein forms insoluble fibrils as well as various soluble oligomers. Although the state involved in disease progression is still a matter of dispute (4–8), it is clear that the population of the different aggregated states is a function of the kinetic processes governing their formation. In particular, the fibril state serves as an important sink of protein mass and may be related to distinct clinical outcomes (9). In addition, understanding the mechanism of β-sheet rich assemblies may also be valuable for the design of biomaterials (2, 10).

Experiments have shown that amyloid β (Aβ) fibrils grow one monomer a time, which has been described by a two-step "Dock and Lock" model (11–14). In this model, an Aβ monomer rapidly adheres to a preformed fibril (dock step), followed by a slow lock step where the unstructured incoming peptide adopts an extended β-conformation (8, 11, 15). However, further experimental dissection of the aggregation process is hindered by numerous metastable substates that interconvert rapidly (3). Molecular dynamics (MD) simulation is a powerful tool that can provide a detailed molecular picture to complement experiments (16–18). However, the experimental elongation rate is roughly one layer per second (19, 20), which is not currently feasible with atomistic simulations. While this time scale is accessible to coarse-grained approaches (18, 21–25), the reduced resolution of these models makes it challenging to distinguish sidechains with similar properties and capture mutational effects.

The long time scale of amyloid aggregation contrasts with native protein folding where secondary structure formation occurs on a submicrosecond time scale. To explain this difference, we presented an analytic theory showing that aggregation kinetics were consistent with an ergodic search of β-sheet alignments (26). By identifying the slowest step in the aggregation process, this theory suggests that high-resolution simulations can be realized by accelerating the alignment search. The proof of concept for this new approach was demonstrated in a paper showing how the alignment search could be captured using a Markov State Model (MSM) defined by the two reaction coordinates identified by the theory (27). That work successfully recapitulated nonadditive effects of hydrophobic mutations but was incomplete in several ways. Key steps in the process, including the initial binding, peptide backbone conformational reconfiguration, and transitions between alignments, were either lumped into "black box" states in the MSM or absorbed into free parameters. Secondly, the implicit solvent model, utilized to allow direct comparison between the MSM and brute force simulations, did not allow a direct comparison of the simulation and experiments. As a result, this "minimalist" treatment lacks quantitative predictive power and failed to provide

Significance

Amyloid fibrils are a type of protein aggregate associated with many neurodegenerative disorders including Alzheimer's disease. The molecular mechanisms of fibril aggregation have been elusive due to the long time scales involved. Here we present a multiscale simulation algorithm that allows us to "fast forward" through the slow parts of the process and focus all-atom sampling time on faster processes. The results reveal that, in contrast to the folding of native proteins, the kinetics of amyloid aggregation is dominated by nonproductive pathways. This multiscale approach does not involve any free parameters and can be extended to describe other protein binding and assembly processes.

Author contributions: Z.J., J.D.S., and J.C. designed research; Z.J. performed research; Z.J. analyzed data; and Z.J., J.D.S., and J.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. R.V.P. is a guest editor invited by the Editorial Board.

Data deposition: The script for Markov State simulation of fibril growth as well as the transition matrix for the wild-type $A\beta_{16-22}$ peptide are available from GitHub (https://github.com/zhiguangjia/MSM-model-for-amyloid-).

[1]To whom correspondence may be addressed. Email: schmit@phys.ksu.edu or jianhanc@umass.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1911153117/-/DCSupplemental.

First published April 28, 2020.

a mechanistic explanation of the curious nonadditive mutation effects.

The present work addresses all of these deficiencies. First, we show how the two-dimensional (2D) MSM can be integrated as the core of a larger MSM describing the full aggregation process. This allows us to compute aggregation rates without free parameters and extensively sample states not included in the original theory. The full MSM shows a detailed mechanism of assembly where weakly bound states serve as a hub connecting the assembly pathway to long-lifetime trapped states. Secondly, due to the highly parallelizable nature of our MSM, we have been able to transition to an explicit solvent model, allowing for direct calculation of aggregation rates. Finally, we explain how the unsatisfying finding that mutation effects are the unpredictable effect of small perturbations is, in fact, the expected outcome within the funnel model of protein folding.

Results

Fibril Elongation Consists of Diffusion, Nonspecific Binding, and β-Alignment Stages. We have developed a computational framework that can systematically explore the search over β-sheet alignments (26, 27). In this framework, the number of hydrogen bonds (H-bond) and alignment (referred to as the "registry") between an incoming peptide and the existing fibril are used as reaction coordinates. The process of exploring registries is then converted to a series of H-bond formation and breakage events. Since H-bond formation and breakage occur on the nanosecond time scale, targeted simulations can be deployed to calculate the rates of these events and measure how they are affected by protein sequence and/or environmental factors. The microscopic rates can be then used to construct MSMs for simulating the overall growth process. This is similar to the application of MSMs for protein folding (28–34). The difference is that in our previous work, the state space was obtained from the analytic theory, while traditional applications of MSMs obtain states from kinetic clustering of simulation trajectories. In the present work, we combine the strengths of the two approaches by using the H-bond state space to sample the slow search over registries and atomistic simulations followed by kinetic clustering to resolve states outside of the H-bonding space.

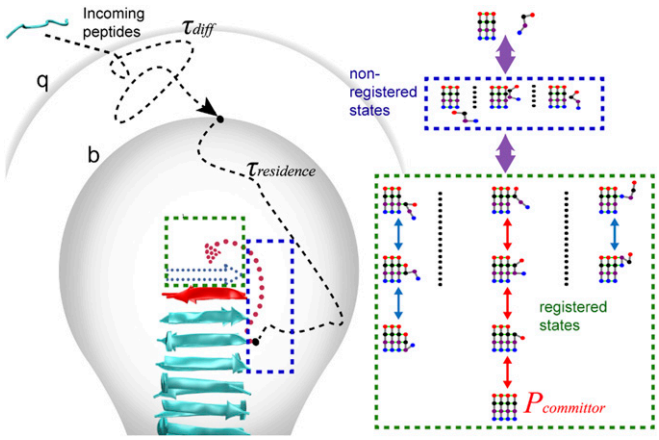

Fig. 1. Schematic illustration of the random walk model of fibril growth with key kinetic parameters required for calculating the growth rate. The incoming peptide migrates from solvent to a separation distance b with a time τ_{diff} (first stage). Then within $\tau_{residence}$, it explores multiple nonregistered (blue box) and registered states (green box) (second and third stage). The peptide is considered to return to the bulk solvent if it reaches the q surface. $P_{committor}$ is the probability of the peptide arriving at the fully bound in-register state (the red path).

Our model is shown schematically in Fig. 1. The elongation process starts from an unstructured monomer in the bulk solvent, which "docks" to a preformed fibril core and "locks" as a new structured layer (11, 12, 15). The process can be divided into three stages. The first stage is a diffusion-controlled process, in which the incoming peptide migrates from solvent to a separation distance b (35, 36). This stage is similar to Brownian dynamics treatments of diffusion-controlled reactions (35, 36). Simulations were started from a monomer at the b surface, and the rates of monomers reaching the fibril surface or an escape distance q were analyzed. If the monomer reaches the q surface, the monomer is considered to have returned to the bulk solvent. The second stage describes nonspecific interactions between the monomer and fibril, defined as those states without either in-register or misregistered backbone H-bonds (nonregistered states; blue box in Fig. 1). This stage ends when the incoming peptide either forms backbone H-bonds with the core peptide, thereby entering the registered stage (third stage) or dissociates from the fibril surface, returning to the unbound state (the q surface). The third stage represents the peptide's exploration of β-strand alignments by forming or breaking H-bonds (green box in Fig. 1). The attempt ends when the molecule either forms a full set of H-bonds in the correct registry (red arrow in Fig. 1, *Right*) or falls off the end of the fibril (returning to nonregistered states). By combining the three stages, our model describes the complete process of fibril growth. The net growth rate can be expressed as

$$k_{\text{growth}} = k_{\text{on}} - k_{\text{off}} = \frac{P_{\text{committor}}}{\tau_{\text{diff}} + \tau_{\text{residence}}} - \frac{1}{\tau_{\text{off}}}, \qquad [1]$$

in which τ_{diff} is the average time for a monomer to diffuse from bulk solvent to the b surface, $\tau_{\text{residence}}$ is the average time required for a molecule entering the b surface to reach either the fully bound state or return to a separation distance q, τ_{off} represents the average time required for a fully bound peptide to reach the dissociated state (q surface), and $P_{\text{committor}}$ is the probability that an incoming peptide starting from the b surface becomes incorporated into the fibril in a fully bound in-register state. Although mathematically similar to the Michaelis–Menten equation, the two rate constants appearing in the first term of Eq. **1** refer to a diffusion step and a reconfiguration step, rather than a two-step reaction that occurs after binding (26, 37).

CHA Mutations Induce Many Small Perturbations in H-Bond Transition Rates. We focus on the fibril growth of the hydrophobic core of Aβ ($A\beta_{16-22}$, $K_{16}LVFFAE_{22}$), which is among the shortest sequences that form fibrils similar to those of full-length Aβ (38–40). All possible register states can be enumerated using a notation similar to our previous work (27) (Fig. 2 and *SI Appendix*, Fig. S2). A large number of 50-ns explicit solvent simulations were deployed to derive H-bond transition rates (*Methods* and *SI Appendix*, Table S1), which enables more realistic rates and allows better description of peptide conformational fluctuations. It turns out that explicit treatment of solvent also necessitates the inclusion of a type of transition between registered states, to account for conformational changes that occur at time scales comparable to those of H-bond formation and breakage (*Methods*). These transitions will also be necessary when extending the current framework to longer peptides.

As summarized in *SI Appendix*, Fig. S3, the effects of CHA mutations are not localized to the mutation site; instead, many H-bond transition kinetics are either increased or decreased. This is similar to our observations in the implicit solvent model (27). Thus, the impact of replacing phenylalanine (PHE) by CHA on fibril growth could not be directly inferred by changes in H-bond transition rates, but should be attributed to an accumulation of small effects during fibril growth.

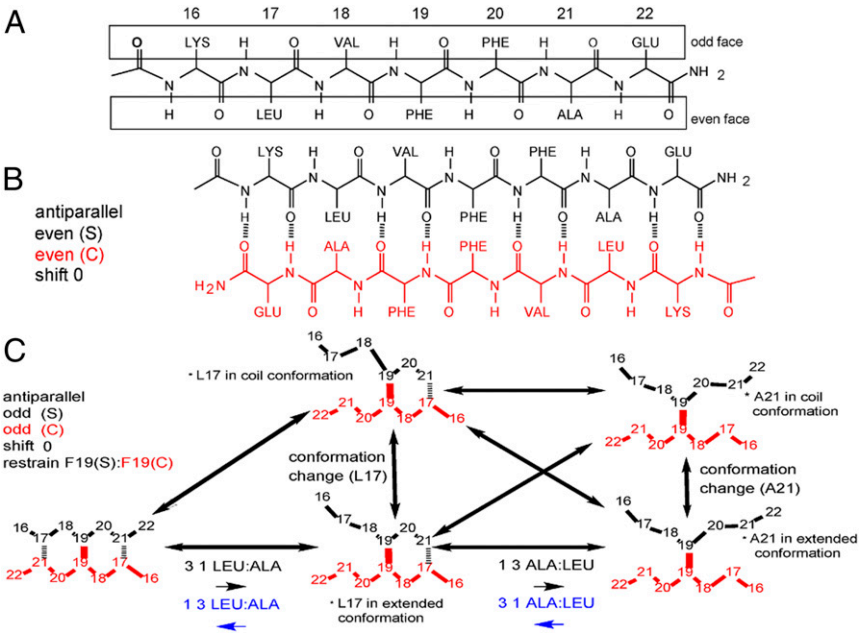

Fig. 2. Backbone H-bonding interactions and states in fibril growth of $A\beta_{16-22}$. (*A*) A preformed $A\beta_{16-22}$ fibril has two faces, odd and even, for templating β-sheet formation of an incoming peptide. (*B*) Illustration of the H-bonding network of the in-register state antiparallel e|e|0. (*C*) Illustration of all transitions of an incoming peptide with the H-bonding pair 19–19 restrained. Note that internal conformational states of the incoming peptide (extended and coil) are explicitly included. ALA, alanine; GLU, glutamine; LEU, leucine; LYS, lysine; VAL, valine.

States without β-Sheet Order Serve as a Kinetic Hub. The nonregistered states do not have an identifiable β-sheet alignment and are not included in our original theory (26). However, our implicit solvent simulations showed that they served as a hub connecting β-sheet registries as well as the disassociated state (27). The implicit solvent treatment did not capture the details of nonregistered states and they were agglomerated into a single "nonspecific state." Explicit solvent simulations revealed that many nonregistered substates exist, and they play distinct roles in mediating fibril growth (*Methods*). For $A\beta_{16-22}$, clustering analysis identified 15 distinct macrostates (Fig. 3), only 3 of which were found close to the docking face of the fibril and connected to registered states. The average transition rates between nonregistered states (~1 ns) are faster than the average formation rate of the first H-bond pair. The transitions between registered states and the three connected nonregistered states occur in ~2 to 6 ns. This suggests that the incoming peptide visits many nonregistered states before forming the first pair of hydrogen bonds. In addition, for a specific nonregistered state, all transitions to registered states can be fitted to the same single-exponential function (*SI Appendix*, Fig. S5), which suggests the transition rate between nonregistered states and the registered states are independent of orientation (antiparallel or parallel) or residue sidechains and only depend on the properties of the nonregistered state (position, peptide conformation, etc.). Thus, a uniform H-bond formation/breakage transition rate was used for each nonregistered state.

MSM Captures the Binding Lifetimes of Different Registries. The final MSM was tested by comparing the calculated lifetimes with results derived directly from unrestrained atomistic simulations for three fully H-bonded antiparallel registries. The results (*SI Appendix*, Fig. S6) show a good agreement between MSM and direct atomistic simulations. Next, we applied the MSM to analyze the effects of PHE-to-CHA mutations on the registry lifetimes. As summarized in *SI Appendix*, Fig. S6 the effects of the single and double CHA mutations are similar on each registry, although the magnitudes of lifetime changes are different. For both mutations, the time of one in-register state "antiparallel e|e|0" is increased and the lifetime for the other in-register state "antiparallel o|o|0" is decreased. Also, three antiparallel misregistered states have longer lifetimes for mutated peptides, and three other misregistered states have shorter lifetimes. As such, the net effects of mutations on the overall fibril growth kinetics is not obvious from the registry lifetimes. Instead, it is necessary to examine the ensemble of MSM trajectories to account for the many effects of the mutations.

Expanded MSM Recapitulates Mutational Effects and Solubility Concentrations. To derive the overall fibril growth kinetics, 10,000 MSM simulations were performed, with 5,000 each on the even and odd faces of the fibril core, to calculate $\tau_{\text{residence}}$ and $P_{\text{committor}}$

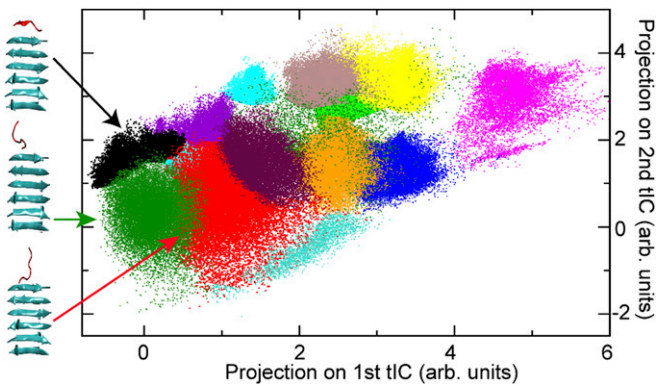

Fig. 3. Nonregistered states identified from kinetic clustering. The simulation trajectories were clustered based on time-lagged independent component analysis (tICA). The states are projected on first two tICAs. The three clusters connected to registered states are colored in black, green, and red. Representative conformations (cluster centers) of the incoming peptide (red) and on the fibril surface (cyan) are shown. Note, as the cluster is based on three-dimensional tICA, some clusters are covered by other clusters and do not show up in the 2D projection.

(Eq. **1**). Each simulation was initiated from the separation distance b and terminated when the incoming strand reached the antiparallel fully H-bonded in-register state or the escape distance q. $P_{committor}$ was calculated as the probability of MSM trajectories reaching the fully H-bonded in-register state, and $\tau_{residence}$ is the average length of all trajectories (Fig. 1*B*). To calculate τ_{off}, another set of 10,000 MSM simulations was initiated from the antiparallel in-register fully bound state and performed until the strand reached the escape distance q. $P_{committor}$ is nearly doubled compared with our previous study with implicit solvent and a single agglomerated nonregistered state (Table 1). This suggests the in-register states are more stable for $A\beta_{16-22}$ in the explicit solvent model. In addition, both $\tau_{residence}$ and τ_{off} are increased by 2 to ∼3 orders of magnitude from the previous study due to the increased residence time of each state (Table 1).

To compare with experimental results, the elongation rate as a function of monomer concentration is estimated. The diffusion time (τ_{diff}; Eq. **1**) is estimated as $\tau_{diff} = 1/k_{coll}\, C_{monomer}$, in which $C_{monomer}$ is concentration of peptide monomer and k_{coll} is collision rate constant. The k_{coll} is defined as $k_{coll} = 4\pi\sigma D_{monomer}$, in which σ is the reaction cross section (b surface in our simulation, 28 Å), and $D_{monomer}$ is the diffusion coefficient (determined from experiments to be 1.7×10^{-6} cm^2/s) (13). The results, plotted in Fig. 4, reveal that the critical concentration (C_c) of peptide monomer from the MSM is ∼8 μM, which is close to the experimental result of ∼33 μM (41). The ability of the current multiscale modeling approach to recapitulate the critical concentration to a factor of ∼4 without any adjustable parameters is noteworthy, particularly considering the comprehensive description of the conformational search process. This also likely reflects significant improvements of explicit solvent force fields made in recent years (42, 43).

To further investigate the influence of mutations, we examine the origin of the change in elongation rate for the mutated peptides. The increased growth rate of the mutants is due to a decrease of $\tau_{residence}$, which in turn, arises from destabilizing two main trap states, "parallel o|o|0" and "parallel e|o|−2" (*SI Appendix*, Table S3). Although the relative free-energy of registered states between the CHA19 and CH1920 double mutation is similar, the double mutation also leads to a decrease in the residence time of in-registered states (*SI Appendix*, Table S2), which leads to a faster dissociation rate (smaller τ_{off}) and a decrease in the growth rate.

Experimental studies have suggested that CHA mutations at the 19 and 20 positions have nonadditive effects on overall aggregation kinetics (the growth rate: wild type [wt] << CHA19 < CHA1920 < CHA20 under high monomer concentration [>40 μM]) (41). However, this result could not be directly explained by thermodynamic effects of the mutations, which appear to have an additive effect (the $\Delta\Delta G$ with respect to wt for CHA19, CHA20, and CHA1920 are 0.2, −0.2, and 0.0 kcal mol^{-1}, respectively). To investigate the influence of mutations, MSM simulations of CHA mutants were performed. In the initial high monomer-concentration condition, both CHA19 and CHA1920 exhibited significant kinetic enhancement in fibril elongation (Fig. 4), in agreement with experiments (41). This corresponds to the initial stage (0 to 2 h) characterized by a fast decrease in the

Table 1. Key kinetic parameters of fiber growth derived from MSM

	Wt	CHA19	CHA1920
$P_{committor}$	0.87 ± 0.08	0.88 ± 0.09	0.90 ± 0.02
$\tau_{residence}$, ns	1,359.0 ± 56.6	546.2 ± 12.2	588.9 ± 1.25
τ_{off}, ns	40,012 ± 515	41,733 ± 836	23,086 ± 522

Data are means ± SD.

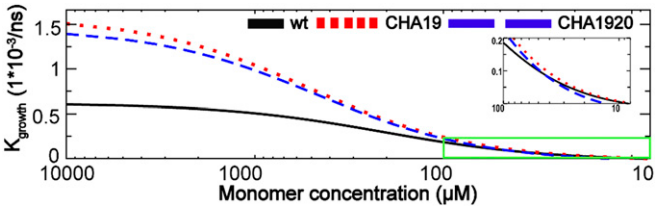

Fig. 4. Predicted net growth rate of $A\beta_{16-22}$ fibril as a function of the peptide concentration. Results for the wt, CHA19, and CHA1920 $A\beta_{16-22}$ are presented as black solid trace, red dotted trace, and blue dash trace, respectively.

monomer concentration. This rapid aggregation reduces the monomer concentration, leading to an increase of τ_{diff}, which, in turn, leads to a decrease of k_{growth}. When the monomer is further depleted, the growth rate approaches zero as the monomer approaches the C_c (41).

Fig. 4 shows a saturation in the growth rate at high concentration as the system transitions from diffusion-limited to reaction-limited. This saturation has been observed in the elongation of insulin fibrils (20). In this regime, new molecules tend to arrive at the fibril ends before the previous molecule has completed the search for the in-register state (26). The nonlinearity in the growth rate as a function of concentration indicates that incorrectly bound molecules inhibit further elongation, presumably by offering a poor surface for binding. An alternative to the saturation behavior is that misregistered molecules become incorporated in the fibril as defects (44). These two outcomes could be distinguished in a simulation by attempting to dock a new protein while the previous one is still misregistered and comparing the residence time to the diffusion time. Experimentally, this could be achieved by measuring the elongation rate (absent nucleation effects; ref. 45) in the concentration regime where Fig. 4 shows nonlinearity.

Discussion

The Aggregation Process Is Not Guided by the Free-Energy Landscape. To investigate the association/dissociation pathways, all MSM trajectories were combined and the relevant free energy of each H-bonding state was plotted. This reveals rough free-energy landscapes with numerous local minima (Fig. 5). Importantly, there is no obvious bias to the landscape to guide the system toward the in-register states. Biased landscapes, often described as the funnel model, are necessary for proteins to navigate the overwhelming conformational space as they fold to their native state (46, 47). Without a bias to guide the search, amyloids must exhaustively search the alignment space in order to match the fibril template. This is possible, albeit slow, because of the relative simplicity of the cross-β structure. The free-energy difference between the in-register substates and misregistered substates are <5 kcal/mol. Indeed, the ratio of misregistered states in amyloid is much higher than the misfolded trapped states observed in natural proteins (48, 49). Although the in-register states are more thermodynamically stable (Fig. 5), to reach these states, an incoming $A\beta_{16-22}$ peptide needs to visit different misregistered states an average of ∼22 times before it forms the fully bound in-register states.

The lack of a funnel landscape is not surprising. Pathological aggregates do not confer a fitness advantage; thus, there is no reason to expect these molecules to have evolved efficient aggregation pathways. The funnel model helps to understand the result of our previous paper that the mutational effects could not be attributed to a few specific states or pathways (27). The dominance of a small number of states or pathways is the expected outcome from an efficient search through a funnel-like landscape. Conversely, in the absence of an efficient pathway, the system will

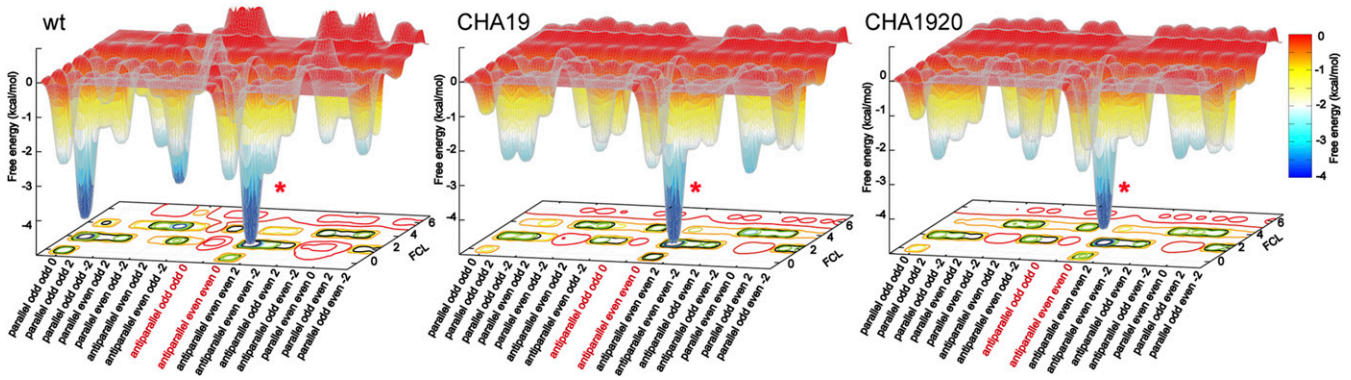

Fig. 5. The free-energy landscape of $A\beta_{16-22}$ fibril growth. The free energy is derived from all MSM trajectories and measured relative to the substate antiparallel e|e|0 with free chain length [FCL] 6 (i.e., with only one H-bond pair formed at either terminus). The two in-registered states antiparallel e|e|0 and antiparallel o|o|0 are highlighted in red. The fully bound antiparallel e|e|0 state is labeled with a red asterisk. The *Inset* shows the detail of the region where the growth curve of wt and mutated peptides are diverged.

randomly sample states throughout the conformational ensemble. This means that the net effect of the mutations will necessarily be the accumulation of perturbations throughout the conformational space. A related question is what the free-energy landscape would look like for a functional amyloid (2). An interesting hypothesis for this scenario arises from recent work showing that functional amyloids form weaker steric zippers (50). While this would not bias the landscape toward more aligned registries, the weaker bonds would smooth out the landscape and facilitate the search by minimizing the depth of the free-energy traps.

Misregistered States Compete with In-Register States. *SI Appendix*, Table S3 shows the free energies for each registry, as estimated from the fraction of time each state was occupied in the MSM. For each molecule, the free energies are measured relative to the lowest occupancy registry. In the wt $A\beta_{16-22}$ peptides, two misregistered states, parallel o|o|0 and parallel e|o|−2, were found to have similar free energy compared with the in-register states (*SI Appendix*, Table S3). We offer two mechanisms to explain the dominance of the antiparallel in-register states over these low energy states in the fibril. First, electrostatic interactions will disfavor parallel states due to the close proximity of both terminal and sidechain charges. The repulsion due to the amino- and carboxyl-terminal charges will be reduced somewhat for molecules at the fibril ends because fluctuations in the binding states will allow the charges to separate. However, when the next molecule attaches, those termini will get pinned down, resulting in the full electrostatic repulsion. This effect is amplified with each additional parallel molecule that is added. The electrostatic contribution from the termini becomes less important with increasing protein length, which explains the parallel structures of full-length Aβ, IAPP, and others. In addition, mutations which weaken these repulsions may allow these nonproductive pathways to dominate fibril assembly. This assumption is consistent with NMR experiments showing that the E22Q mutation leads to parallel fibrils (51). Interestingly, this mutation is the same as the Dutch mutant of full-length Aβ peptide, which enhances the aggregation propensity (52).

Secondly, while misregistered states, like parallel e|o|−2, bind with high affinity, they may offer a poor template for subsequent fibril growth. An out-of-register shift will result in overhanging amino acids that will bind poorly to subsequent molecules due to the conformational entropy penalty of immobilizing the flexible residues (53). However, under conditions conducive to rapid growth (e.g., when the monomer concentration is high), these registry defects may be trapped within the growing fibril by subsequent monomer addition (44).

MSM Trajectories Show a Molecular Mechanism of Amyloid Formation. For further insight, we examine a representative trajectory in Fig. 6*A*, which starts from the dissociated state, reaches the fully bound in-register state, and eventually returns to the dissociated state. In this trajectory, the locking step of the Dock and Lock model occurs in the initial 4 μs. This period is characterized by many rapid transitions between nonregistered states (cyan dots) separated by long periods in misregistered β-sheet states (yellow). The residence time in this stage is dominated by misregistered states (>98%). The end of the locking step (4 μs) begins a long period of stability marked by only small fluctuations from the fully bound state. However, eventually (~7 μs), there is a fluctuation large enough to remove it from the in-register state, and the molecule finally dissociates. This trajectory shows many events where the molecule briefly detaches from the fibril (red dots) before reattaching. These multiple attempts at finding the in-register state explain the large values of $P_{\text{committor}}$ (Table 1) and show why molecules reaching the docked state are much more likely to proceed to the locked states than fail to incorporate.

The overall mechanism is depicted schematically in Fig. 6*B*. The assembly pathway is highlighted by the purple arrow. This pathway includes the diffusion step, initial contact with the fibril in the non–β-sheet state, and the formation of an in-register β-sheet. However, the molecule spends less than 0.1% of its time on this pathway, instead spending most of the time in off-pathway traps. Of these, the most significant, by far, are the misregistered β-sheets (yellow in Fig. 6). These states are deep free-energy wells (Fig. 5) because of the translational symmetry of the peptide backbone. These traps become deeper for increasing peptide length, leading to exponential increases in the aggregation time (45). These long trap lifetimes have been previously compared with glassy systems (54). While these off-pathway states slow fibril growth, they also inhibit dissolution by binding departing peptides (~200 times on average in a single dissociation MSM trajectory). The resulting long τ_{off} is consistent with the experimental result that the locking stage is nearly irreversible (11). The off-pathway states without β-sheet contact, in contrast, are an insignificant contribution to the overall aggregation time (dark blue in Fig. 6). However, they are important in keeping the molecule in close proximity to the fibril. Finally, the hub states (teal in Fig. 6) play a crucial role in facilitating assembly. These states are similar to the non-β traps morphologically, but play a special role because they are close enough to the fibril end to allow direction conversion to and from β-sheet states.

Based on these results, we can speculate on how the assembly mechanism might change for longer peptides such as $A\beta_{1-42}$. A

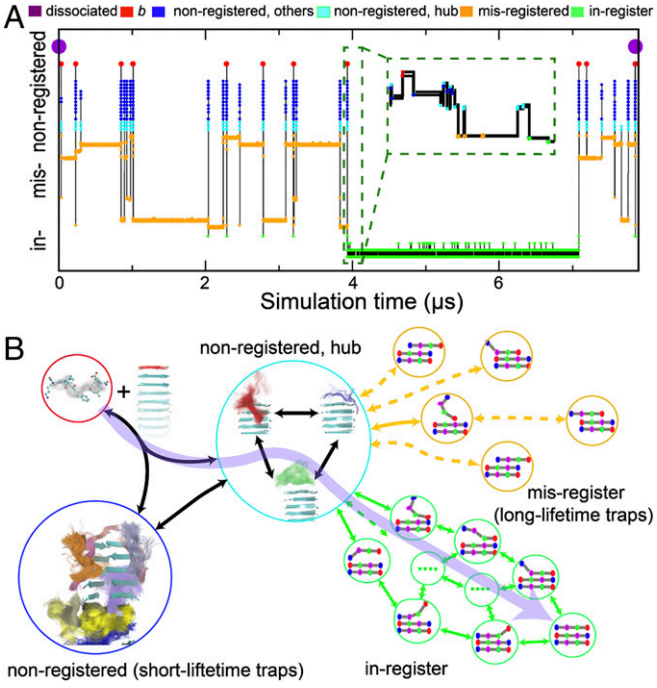

Fig. 6. Overall mechanism of $A\beta_{16-22}$ fibril growth. (*A*) A sample MSM trajectory illustrating the reversible templating transition of an incoming peptide. The trajectory was initiated from a fully dissociated state (marked by the purple dots) and continued until the peptide returned to the dissociated state. The *y* axis shows the MSM states sampled during the simulation. The MSM states were consolidated based on the registry and total free chain length for clarity (also see Fig. 4). The *Inset* shows the detail of the transition from nonregistered states (cyan) to the in-register (green) fibril state (near 3.93 μs). (*B*) Schematic of the fibril elongation mechanism. The aggregation pathway (purple arrow) takes proteins from the soluble state to in-register β-sheet states (green circles), but it accounts for a small fraction of the aggregation time. Instead, molecules spend >98% of the time trapped in misregistered β-sheet states (orange circles). States without β-sheet structure present another off-pathway trap (blue circle), although a few non-β states serve as a hub (cyan circle) connecting the trap states with the aggregation pathway. The assembly pathway is highlighted by the transparent purple arrow.

major difference with longer molecules is their ability to fold in the solution state. One line of thought is that adoption of the aggregated conformation in the solution state promotes aggregation (55–58). Based on the mechanism observed in our simulations, we believe this hypothesis is unlikely. This is because the formation of tertiary contacts would inhibit the conformational flexibility needed to sample the registry space; trapping segments of the backbone in β-conformation would prevent it from peeling off the template one residue at a time. This means that much larger energy fluctuations would be required to rectify alignment errors. Therefore, formation of ordered fibrils is most likely from peptides that are disordered in the monomer state. However, adoption of the aggregated state will be beneficial for fibril nucleation (53).

Conclusion

Functional proteins have evolved to develop efficient folding pathways where a funnel-shaped free-energy landscape reduces the time spent exploring nonproductive states. This evolutionary pressure does not exist for nonfunctional states like pathological aggregates. In these cases, the free-energy landscape would more closely resemble a golf course (47) or an inverted funnel (44). Without the guidance of a biased landscape, the conformational search will resemble a random search over numerous H-bonded and non–H-bonded binding states. We have developed an analytic theory-motivated, multiscale simulation strategy to derive a comprehensive description of productive and nonproductive pathways in the fibril growth of $A\beta_{16-22}$. Without any free parameters, these descriptions quantitatively reproduce the critical fibril growth concentrations and recapitulate the effects and nonadditivity of CHA19 and CHA20 mutations on the growth kinetics. The results reveal that, instead of finding a few dominant pathways or intermediate states, the effects of mutations emerge from the accumulation of small perturbations over the ensemble of partially bound states. This recognition poses challenges for computational methods, which must sample both productive and nonproductive pathways to generate meaningful predictions. We note that majority of modeling treatments of fibril growth to date use pure MD simulations (56–58). They are limited in the conformational search processes that can be explored, even with coarse-grained treatment and/or enhanced sampling techniques. The current work demonstrates how insights from analytic theory can be utilized to meet this challenge by inspiring sampling techniques that can access new time scales and provide a more complete understanding of fibril growth.

Methods

A total of 719 μs of explicit solvent simulations in the CHARMM36m all-atom force field (59, 60) was performed to derive the microscopic kinetic parameters for transition among all H-bonded and nonspecific bound states. See *SI Appendix, Supporting Text* for detailed simulation and analysis procedures.

Data Availability Statement. The script for Markov State simulation of fibril growth as well as the transition matrix for the wt $A\beta_{16-22}$ peptide are available from GitHub (https://github.com/zhiguangjia/MSM-model-for-amyloid-). The CHARMM scripts for atomistic simulation, analysis, clustering, and kinetic analysis are available upon request.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grant GM107487. Computing for this project was performed on the Beocat Research Cluster at Kansas State University and the Pikes cluster housed in the Massachusetts Green High-Performance Computing Center.

1. A. Morriss-Andrews, J.-E. Shea, Computational studies of protein aggregation: Methods and applications. *Annu. Rev. Phys. Chem.* **66**, 643–666 (2015).
2. D. M. Fowler, A. V. Koulov, W. E. Balch, J. W. Kelly, Functional amyloid–From bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224 (2007).
3. T. C. T. Michaels *et al.*, Chemical kinetics for bridging molecular mechanisms and macroscopic measurements of amyloid fibril formation. *Annu. Rev. Phys. Chem.* **69**, 273–298 (2018).
4. C. Haass, D. J. Selkoe, Soluble protein oligomers in neurodegeneration: Lessons from the Alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol.* **8**, 101–112 (2007).
5. J. Hardy, D. J. Selkoe, The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).
6. M. D. Kirkitadze, G. Bitan, D. B. Teplow, Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: The emerging role of oligomeric assemblies. *J. Neurosci. Res.* **69**, 567–577 (2002).
7. J. R. Silveira *et al.*, The most infectious prion protein particles. *Nature* **437**, 257–261 (2005).
8. A. Morriss-Andrews, J.-E. Shea, Simulations of protein aggregation: Insights from atomistic and Coarse-Grained models. *J. Phys. Chem. Lett.* **5**, 1899–1908 (2014).
9. W. Qiang, W.-M. Yau, J.-X. Lu, J. Collinge, R. Tycko, Structural variation in amyloid-β fibrils from Alzheimer's disease clinical subtypes. *Nature* **541**, 217–221 (2017).
10. T. P. J. Knowles, R. Mezzenga, Amyloid fibrils as building blocks for natural and artificial functional materials. *Adv. Mater.* **28**, 6546–6561 (2016).
11. W. P. Esler *et al.*, Alzheimer's disease amyloid propagation by a template-dependent dock-lock mechanism. *Biochemistry* **39**, 6288–6295 (2000).
12. E. P. O'Brien, Y. Okamoto, J. E. Straub, B. R. Brooks, D. Thirumalai, Thermodynamic perspective on the dock-lock growth mechanism of amyloid fibrils. *J. Phys. Chem. B* **113**, 14421–14430 (2009).
13. T. Ban, K. Yamaguchi, Y. Goto, Direct observation of amyloid fibril growth, propagation, and adaptation. *Acc. Chem. Res.* **39**, 663–670 (2006).
14. S. R. Collins, A. Douglass, R. D. Vale, J. S. Weissman, Mechanism of prion propagation: Amyloid growth occurs by monomer addition. *PLoS Biol.* **2**, e321 (2004).

15. P. H. Nguyen, M. S. Li, G. Stock, J. E. Straub, D. Thirumalai, Monomer adds to preformed structured oligomers of Abeta-peptides by a two-stage dock-lock mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 111–116 (2007).
16. B. Ma, R. Nussinov, Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Curr. Opin. Chem. Biol.* **10**, 445–452 (2006).
17. J. E. Straub, D. Thirumalai, Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annu. Rev. Phys. Chem.* **62**, 437–463 (2011).
18. C. Wu, J.-E. Shea, Coarse-grained models for protein aggregation. *Curr. Opin. Struct. Biol.* **21**, 209–220 (2011).
19. T. Ban *et al*., Direct observation of Abeta amyloid fibril growth and inhibition. *J. Mol. Biol.* **344**, 757–767 (2004).
20. T. P. J. Knowles *et al*., Kinetics and thermodynamics of amyloid formation from direct measurements of fluctuations in fibril mass. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10016–10021 (2007).
21. H. D. Nguyen, C. K. Hall, Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16180–16185 (2004).
22. S. Auer, F. Meersman, C. M. Dobson, M. Vendruscolo, A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates. *PLOS Comput. Biol.* **4**, e1000222 (2008).
23. P. Ricchiuto, A. V. Brukhno, S. Auer, Protein aggregation: Kinetics versus thermodynamics. *J. Phys. Chem. B* **116**, 5384–5390 (2012).
24. M. Cheon, I. Chang, C. K. Hall, Spontaneous formation of twisted Aβ(16-22) fibrils in large-scale molecular-dynamics simulations. *Biophys. J.* **101**, 2493–2501 (2011).
25. M. Cheon, I. Chang, C. K. Hall, Extending the PRIME model for protein aggregation to all 20 amino acids. *Proteins* **78**, 2950–2960 (2010).
26. J. D. Schmit, Kinetic theory of amyloid fibril templating. *J. Chem. Phys.* **138**, 185102 (2013).
27. Z. Jia, A. Beugelsdijk, J. Chen, J. D. Schmit, The Levinthal problem in amyloid aggregation: Sampling of a flat reaction space. *J. Phys. Chem. B* **121**, 1576–1586 (2017).
28. M. Schor, A. S. J. S. Mey, F. Noé, C. E. MacPhee, Shedding light on the Dock–Lock mechanism in amyloid fibril growth using Markov State Models. *J. Phys. Chem. Lett.* **6**, 1076–1081 (2015).
29. B. Barz, D. J. Wales, B. Strodel, A kinetic approach to the sequence-aggregation relationship in disease-related protein assembly. *J. Phys. Chem. B* **118**, 1003–1011 (2014).
30. T. J. Lane, D. Shukla, K. A. Beauchamp, V. S. Pande, To milliseconds and beyond: Challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **23**, 58–65 (2013).
31. M. Senne, B. Trendelkamp-Schroer, A. S. J. S. Mey, C. Schütte, F. Noé, EMMA: A software package for Markov model building and analysis. *J. Chem. Theory Comput.* **8**, 2223–2238 (2012).
32. V. S. Pande, K. Beauchamp, G. R. Bowman, Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**, 99–105 (2010).
33. J. D. Chodera, F. Noé, Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).
34. G. R. Bowman, X. Huang, V. S. Pande, Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **49**, 197–201 (2009).
35. S. H. Northrup, S. A. Allison, J. A. McCammon, Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* **80**, 1517–1524 (1984).
36. B. A. Luty, J. A. McCammon, H. X. Zhou, Diffusive reaction rates from Brownian dynamics simulations: Replacing the outer cutoff surface by an analytical treatment. *J. Chem. Phys.* **97**, 5682–5686 (1992).
37. A. K. Buell *et al*., Frequency factors in a landscape model of filamentous protein aggregation. *Phys. Rev. Lett.* **104**, 228101 (2010).
38. J. J. Balbach *et al*., Amyloid fibril formation by A β 16-22, a seven-residue fragment of the Alzheimer's β-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* **39**, 13748–13759 (2000).
39. A. K. Mehta *et al*., Facial symmetry in protein self-assembly. *J. Am. Chem. Soc.* **130**, 9829–9835 (2008).
40. A. T. Petkova *et al*., Solid state NMR reveals a pH-dependent antiparallel β-sheet registry in fibrils formed by a β-amyloid peptide. *J. Mol. Biol.* **335**, 247–260 (2004).
41. F. T. Senguen *et al*., Probing aromatic, hydrophobic, and steric effects on the self-assembly of an amyloid-β fragment peptide. *Mol. Biosyst.* **7**, 486–496 (2011).
42. J. Huang, A. D. MacKerell, Jr, Force field development and simulations of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **48**, 40–48 (2018).
43. K. Vanommeslaeghe, A. D. MacKerell, Jr, CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* **1850**, 861–871 (2015).
44. C. Huang, E. Ghanati, J. D. Schmit, Theory of sequence effects in amyloid aggregation. *J. Phys. Chem. B* **122**, 5567–5578 (2018).
45. S. I. A. Cohen *et al*., Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9758–9763 (2013).
46. K. Okazaki, N. Koga, S. Takada, J. N. Onuchic, P. G. Wolynes, Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11844–11849 (2006).
47. K. A. Dill, H. S. Chan, From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
48. S. Piana, K. Lindorff-Larsen, D. E. Shaw, Atomistic description of the folding of a dimeric protein. *J. Phys. Chem. B* **117**, 12935–12942 (2013).
49. L. Sborgi *et al*., Interaction networks in protein folding via atomic-resolution experiments and long-time-scale molecular dynamics simulations. *J. Am. Chem. Soc.* **137**, 6506–6516 (2015).
50. M. P. Hughes *et al*., Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* **359**, 698–701 (2018).
51. C. Liang *et al*., Kinetic intermediates in amyloid assembly. *J. Am. Chem. Soc.* **136**, 15146–15149 (2014).
52. S. G. van Duinen *et al*., Hereditary cerebral hemorrhage with amyloidosis in patients of Dutch origin is related to Alzheimer disease. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5991–5994 (1987).
53. L. Zhang, J. D. Schmit, Pseudo-one-dimensional nucleation in dilute polymer solutions. *Phys. Rev. E* **93**, 060401 (2016).
54. J. K. Weber, R. L. Jack, C. R. Schwantes, V. S. Pande, Dynamical phase transitions reveal amyloid-like states on protein folding landscapes. *Biophys. J.* **107**, 974–982 (2014).
55. M. S. Li, D. K. Klimov, J. E. Straub, D. Thirumalai, Probing the mechanisms of fibril formation using lattice models. *J. Chem. Phys.* **129**, 175101 (2008).
56. J. E. Straub, D. Thirumalai, Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annu. Rev. Phys. Chem.* 62, 437–463 (2011).
57. J. Nasica-Labouze *et al*., Amyloid β protein and Alzheimer's disease: When computer simulations complement experimental studies. *Chem. Rev.* **115**, 3518–3563 (2015).
58. I. M. Ilie, A. Caflisch, Simulation studies of amyloidogenic polypeptides and their aggregates. *Chem. Rev.* **119**, 6956–6993 (2019).
59. J. Huang, A. D. MacKerell, Jr, CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
60. J. Huang *et al*., CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).