# Differences in the free energies between the excited states of A$\beta$40 and A$\beta$42 monomers encode their aggregation propensities

Debayan Chakraborty[a], John E. Straub[b], and D. Thirumalai[a,1]

[a]Department of Chemistry, The University of Texas at Austin, Austin, TX 78712; and [b]Department of Chemistry, Boston University, Boston, MA 02215

Edited by Ken A. Dill, Stony Brook University, Stony Brook, NY, and approved June 26, 2020 (received for review February 10, 2020)

**The early events in the aggregation of the intrinsically disordered peptide, amyloid-$\beta$ (A$\beta$), involve transitions from the disordered free energy ground state to assembly-competent states. Are the fingerprints of order found in the amyloid fibrils encoded in the conformations that the monomers access at equilibrium? If so, could the enhanced aggregation rate of A$\beta$42 compared to A$\beta$40 be rationalized from the sparsely populated high free energy states of the monomers? Here, we answer these questions in the affirmative using coarse-grained simulations of the self-organized polymer–intrinsically disordered protein (SOP-IDP) model of A$\beta$40 and A$\beta$42. Although both the peptides have practically identical ensemble-averaged properties, characteristic of random coils (RCs), the conformational ensembles of the two monomers exhibit sequence-specific heterogeneity. Hierarchical clustering of conformations reveals that both the peptides populate high free energy aggregation-prone (*N**) states, which resemble the monomers in the fibril structure. The free energy gap between the ground (RC) and the *N** states of A$\beta$42 peptide is smaller than that for A$\beta$40. By relating the populations of excited states of the two peptides to the fibril formation time scales using an empirical formula, we explain nearly quantitatively the faster aggregation rate of A$\beta$42 relative to A$\beta$40. The *N** concept accounts for fibril polymorphs, leading to the prediction that the less stable *N** state of A$\beta$42, encoding for the U-bend fibril, should form earlier than the structure with the S-bend topology, which is in accord with Ostwald's rule rationalizing crystal polymorph formation.**

protein aggregation | Alzheimer's disease | free energy gap | Ostwald's rule | polymorphism

Alzheimer's disease (AD), a neurodegenerative disorder affecting a large proportion of the world population, is associated with the gradual accumulation of insoluble plaques in various parts of the brain (1–5). The fibrillar structures of the amyloid-$\beta$ (A$\beta$) peptides, like other amyloid aggregates, have a characteristic cross-$\beta$ architecture (6–8). The fibrils form as a result of aggregation of the 39- to 43-residue-long A$\beta$ peptides. In vivo, A$\beta$ is produced by the proteolytic cleavage of the amyloid precursor protein (APP) by $\gamma$-secretases, with A$\beta$40 and A$\beta$42 being the major products (9, 10). The fibrils exhibit a high degree of polymorphism, even though globally they all have the characteristic $\beta$-sheet architecture. The link between polymorphic structures, which could depend on the way they are generated, and the extent of neurotoxicity is a topic of continuing interest (1, 11). In addition to AD, several other neurodegenerative diseases, such as Parkinson's and Huntington's disease, are also the result of abnormal in vivo protein aggregation. Various studies (12, 13) have proposed that these diseases could share several common themes with AD, including similar fibril morphologies, mechanism of fibril formation, and the association of toxicity with oligomers. To obtain a more quantitative understanding of these general principles, it is crucial to understand the key steps involved in amyloid aggregation at the molecular level.

The cascade of events that describe the conversion of monomers to fibrils is complex (14–16). Because of the transient nature of species that form during the early stages (before a critical size nucleus), it is difficult to describe their structural details using experiments alone. A complete understanding requires describing the conformational changes that occur during the aggregation process starting from the monomer. Toward this end, conformational dynamics of A$\beta$ monomers have been extensively studied, using both experiments (17–19) and computer simulations (15, 20–25). The structural details of the predicted ensembles vary widely among the different studies. Some of the differences in the experiments could be due to differing external conditions, such as variations in pH, temperature, or salt concentration. Molecular dynamics simulations reveal that the populations of different secondary and tertiary structural elements are strongly dependent on the details of the force field and sampling strategies (26–29). However, based on recent NMR (17, 18) and Förster resonance energy transfer (FRET) experiments (30), a general consensus seems to be emerging that A$\beta$ monomers behave like many other intrinsically disordered proteins (IDPs) that are highly disordered, adopting random coil (RC) structures at room temperature and neutral pH. The structures are devoid of any persistent secondary structures ($\alpha$-helices and $\beta$-sheets). In other words, it appears that the propensity to form amyloid fibrils cannot be discerned

## Significance

**Alzheimer's disease (AD) is caused by the accumulation of pathological aggregates, primarily composed of the A$\beta$40 and A$\beta$42 isoforms. Experiments have shown that A$\beta$42 is more aggregation prone compared to A$\beta$40. However, the molecular origin of this apparent anomaly remains elusive. Here, we provide a microscopic basis for the different aggregation rates in terms of the distinct populations of high free energy excited fibril-like states (*N**) that are encoded in the monomer spectra. The *N** theory explains the emergence of fibril polymorphs and predicts the relative kinetic stabilities of A$\beta$42 fibrils using Ostwald's rule of stages. Our work shows that sequence-specific conformational heterogeneity of the monomer ensembles provides important cues for understanding protein aggregation.**

Author contributions: D.C., J.E.S., and D.T. designed research; D.C. performed research; D.C. and D.T. contributed new reagents/analytic tools; D.C. and D.T. analyzed data; and D.C., J.E.S., and D.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



Data deposition: The dataset and the analysis scripts for this paper are hosted by Zenodo (https://zenodo.org/record/3891924#.XwxyqJNKgY0).

[1]To whom correspondence may be addressed. Email: dave.thirumalai@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2002570117/-/DCSupplemental.

when only the average properties of the monomer ensemble are examined.

The natural question is whether there are any connections between the conformational heterogeneity of the monomer ensemble in solution and the eventual structures adopted in the fibril state. More precisely, at what stage of A$\beta$ aggregation does one observe a transition from the RC conformation to structures with considerable $\beta$ content? In a series of papers, we (31–33) showed that the structures that have a high propensity to aggregate are encoded as excitations in the monomer free energy spectrum. Subsequent studies (34–36) have confirmed our findings. More importantly, such conformations have some of the structural features of the monomers in the fibril. We first elucidated this concept using atomically detailed simulations of a 26-residue fragment of the A$\beta$ peptide using molecular dynamics simulations (31). Subsequently, we illustrated the consequences quantitatively using lattice models (32), for which precise computations could be carried out by exploring the entire sequence space. Simulations based on the lattice models also showed empirically that the time scales of fibril formation ($\tau_{fib}$) could be linked to the population ($p_{N*}$) of the aggregation-prone species (referred to from now on as $N^*$ states) (32, 36). The structures in the $N^*$ ensembles (there are more than one) are difficult to characterize using standard experimental techniques because they are high free energy states that are only sparsely populated. In this context computations are useful.

For A$\beta$40 and A$\beta$42, the two most prevalent isoforms in vivo, the question of whether there is a link between the underlying heterogeneity (and by inference, $p_{N*}$) of the monomeric ensembles and their aggregation propensities is pertinent. This is because A$\beta$42 aggregates faster and is the major constituent of amyloid plaques (37–39), despite being present in relatively lower abundance than A$\beta$40 (40). Several studies have attributed the elevated pathogenicity of A$\beta$42 to the pronounced structuring near the C terminus, which is mediated by the two extra residues (ILE41 and ALA42) (41–43). Recent A$\beta$42 fibril structures determined using solution-state NMR (44, 45) and cryo-electron microscopy (EM) (46) seem to corroborate this viewpoint. Previous work based on all-atom studies (29, 30, 42, 47, 48), and simulations using coarse-grained models (49–51) further predict that the two isoforms could be associated with different populations of the metastable states and allude to a sequence-specific conformational heterogeneity that could be a key determinant of the contrasting aggregation propensities.

Here, we report results from simulations using a modified version of a recently developed highly accurate coarse-grained (CG) model for intrinsically disordered proteins (termed as self-organized polymer-intrinsically disordered protein [SOP-IDP] model) (52) to characterize the conformational ensembles of A$\beta$40 and A$\beta$42. The major purpose is to illustrate the theory that the differences in the extent of population of the sparsely populated $N^*$ states in the monomeric ensembles of the A$\beta$ peptides quantitatively explain the relative fibril formation rates of A$\beta$42 and A$\beta$40. We show that despite having practically identical ensemble-averaged values of several experimental observables under physiological conditions, the underlying conformational heterogeneities of the two peptides are distinct. In particular, the termini display a different extent of residual structuring in the two peptides. Most importantly, A$\beta$42 has a higher population ($p_{N*}$) of aggregation-prone (excited) states that display many of the characteristic structural features found in fibrils. The difference in the equilibrium populations of $N^*$ states between A$\beta$42 and A$\beta$40 when linked to kinetics rationalizes the nearly one order of magnitude faster fibril formation rate of A$\beta$42 relative to A$\beta$40. Our study, therefore, provides a crucial link between aggregation propensity and sequence-specific conformational heterogeneity in the monomer. In other words, the spectrum of excited, but sparsely populated $N^*$ states sampled at equilibrium is a harbinger of protein aggregation.

## Results

**Sizes and Polymeric Features of A$\beta$ Peptides Are Nearly Identical.** The distributions of the radius of gyration, $R_g$, and end-to-end distance, $R_{ee}$, both of which describe the global dimensions of the A$\beta$ monomers, are similar (Fig. 1 and Table 1). The small differences are likely due the longer length of A$\beta$42. Our estimates for $\langle R_g \rangle$ and $\langle R_{ee} \rangle$ are in good agreement with the values reported in two recent studies (28, 30). The distributions of the scaled end-to-end distance, $x$ ($R_{ee}/\langle R_{ee}^2 \rangle^{1/2}$), shown in Fig. 1*C*, for both the A$\beta$ monomers deviate from the well-established theoretical results for a Gaussian chain and a self-avoiding random walk (SAW) (see *SI Appendix* for further details). The departure from standard homopolymer theories is the first hint that sequence-specific effects as well as the inherent polyampholyte-like characteristic of the A$\beta$ peptides need to be accounted for when describing their conformational ensembles.

Besides $R_g$ and $R_{ee}$, the hydrodynamic radius $R_h$ is another measure of polymer size, which can be estimated using dynamic light scattering (DLS), NMR, and fluorescence correlation spectroscopy (FCS) experiments (53, 54). The distributions of $R_h$ for the A$\beta$40 and A$\beta$42 also overlap to a large extent (Fig. 1*B*) and yield similar ensemble averages (Table 1). For both the sequences, the $\langle R_h \rangle$ values are in excellent agreement with those (1.5 to 1.7 nm) measured in NMR (53) and FCS experiments (54). The ratio $R_h/R_g$ for both A$\beta$40 and A$\beta$42 is $\approx$0.9, which implies that neither peptide behaves as an RC or a Gaussian chain, for which the ratios should be 0.640 and 0.665, respectively (55).

A recent study (30) reported the global dimensions and conformational dynamics of A$\beta$ monomers using single-molecule experiments, thus complementing previous findings from NMR and circular dichroism (CD) experiments. The FRET histograms computed from our simulations, using Eq. **7** (see *SI Appendix*), are shown in Fig. 1*D*. The average FRET efficiencies, $\langle E \rangle$ for both the peptides (Table 1) are in excellent agreement with the experimental estimates ($\approx$0.6) at zero denaturant concentration (30). Unlike in the FRET experiment, however, we also observe a peak at $\langle E \rangle \approx 1$. This discrepancy is not unexpected and could be attributed to the small size of the peptides, as well as the large Förster radius (5.2 nm) of the Alexa 488 and 647 dye pair, which according to Eq. **7** (see *SI Appendix*), would result in $\langle E \rangle \approx 1$ for $R_{ee}$ less than 4.0 nm. Similar observations have been made in two previous studies (28, 30). From all-atom simulations, Meng et al. (30) noted that although the A$\beta$ monomers were characterized by largely overlapping distributions for $R_g$, $R_{ee}$, and $\langle E \rangle$, the $R_{ee}$ distribution of A$\beta$42 exhibited a minor peak at low end-to-end distances, corresponding to a subpopulation of conformations with long-range contacts. However, no such feature is discernible from our computed probability distributions (Fig. 1). The good agreement between our simulations and experiment for the observables in Table 1 attests to the accuracy of the SOP-IDP model in predicting the values of measurable quantities.

**A$\beta$ Ensembles Are Disordered.** For both the sequences, the inter-residue distances, $R_{ij}$, as a function of sequence separation, $|i-j|$, adhere to the Flory scaling law: $R_{ij} = R_0|i-j|^{\nu}$, with $R_0$ = 0.47 nm, and the scaling exponent, $\nu$ = 0.58 (Fig. 2; see *SI Appendix* for further details). The free energy landscapes projected onto the first and second principal components, based on the C$_\alpha$–C$_\alpha$ distances and $R_g$ and $R_{ee}$ coordinates (*SI Appendix*, Figs. S3 and S4), are featureless, further exemplifying the intrinsically disordered nature, as well as the underlying conformational heterogeneity of the A$\beta$ monomer ensembles.

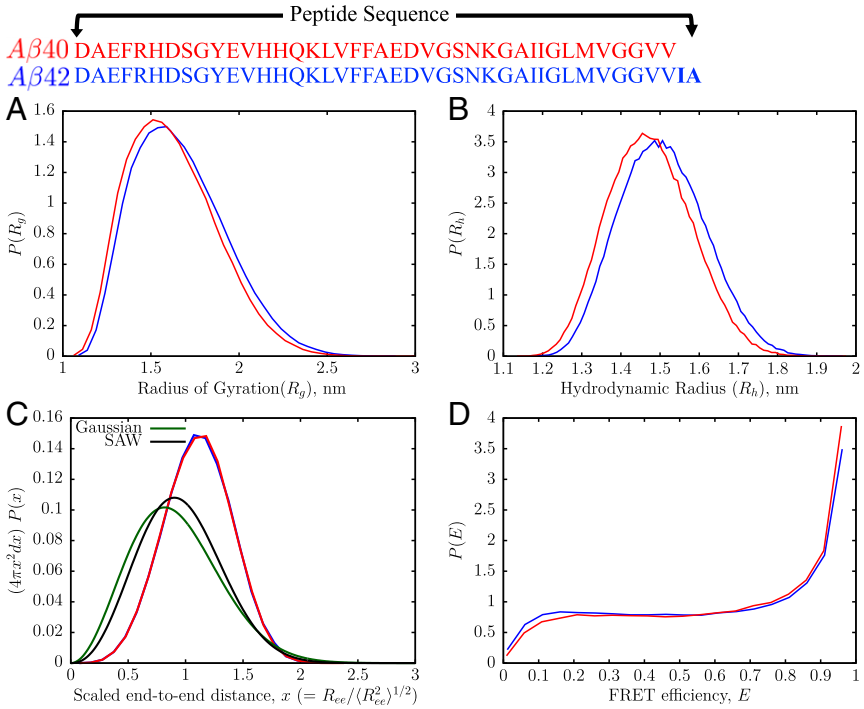


**Fig. 1.** The sequences of A$\beta$40 and A$\beta$42 using one-letter code are displayed at *Top*. Ensemble characteristics of A$\beta$40 and A$\beta$42 are depicted as red and blue curves, respectively. (*A*) Distribution of the radius of gyration, $R_g$. (*B*) Distribution of hydrodynamic radius, $R_h$. (*C*) Distribution of end-to-end distance, $R_{ee}$, scaled by $\langle R_{ee}^2 \rangle^{1/2}$. To compare with the results for a Gaussian chain and a SAW, we show $4\pi x^2 P(x)dx$ where $x = R_{ee}/\langle R_{ee}^2 \rangle^{1/2}$. (*D*) Distribution of FRET efficiencies. The corresponding mean values are given in Table 1.

There are no discernible long-range contacts in the ensemble-averaged contact maps for A$\beta$40 and A$\beta$42 (*SI Appendix*, Fig. S5). However, from a zoomed-in view of the difference map, $\delta p_{ij}$, it is apparent that a small but detectable population of substates (masked in the ensemble-averaged contact map), which have a tendency to form residual structure at the C terminus, is present in the conformational ensemble of A$\beta$42 (Fig. 2). Taken together these results show that A$\beta$ monomers do not adopt stable well-defined secondary or tertiary structures at equilibrium, but rather dynamically interconvert among a menagerie of conformations.

**A$\beta$ Monomers Have Negligible Residual Secondary Structure.** The secondary structure profiles associated with the A$\beta$40 and A$\beta$42 monomers are shown in Fig. 3. Both the peptides have very low preferences for structural elements, such as $\beta$-strand or helices. All of the residues display a very high propensity (>97%) to form turns or irregular coils. Our observations are consistent with NMR experiments (17, 18), as well as all-atom simulations (28, 30), which predict that A$\beta$ monomers populate mostly RC-like structures in equilibrium. Overall, we observe a large overlap between the turn profiles for A$\beta$40 and A$\beta$42. In particular, the elevated turn propensities near residues 7 to 11, as well as 23 to 27 (which include the well-known valine-glycine-serine-asparagine [VGSN] turn), are consistent with the NMR experiment (18). Nonetheless, there are some subtle differences in the propensity of forming residual structure in some parts of the sequence, which are worth noting.

A$\beta$40 has a more structured N terminus (residues ≈4 to 6), as indicated by the marginally higher propensity to form $\beta$-sheets and $\alpha$-helices in this region. In contrast, A$\beta$42 has a slightly more ordered C terminus (≈38 to 42), with the residues in this region displaying a somewhat higher probability of forming both helices and $\beta$-sheets and a pronounced bend near residues 36 to 37. The contrasting structural preferences at the two termini, particularly the enhanced ordering at the C terminus of A$\beta$42, have also been noted in previous all-atom (26–28, 42, 56) and coarse-grained simulations (49, 57). Interestingly, in both the peptides, residues near ≈13 to 20, which include the central hydrophobic core (CHC), show an enhanced probability to form helical domains. As argued in previous experimental (58) and simulation studies (23, 59, 60), such residual structure near the CHC could play a key role in mediating the early events in amyloid fibrillogenesis. We also find that in both the monomers, residues 29 to 36 tend to form secondary structures with higher probability compared to the rest of the sequence. This feature is also

**Table 1. The ensemble-averaged values of different observables for A$\beta$40 and A$\beta$42**

| Sequence | $\langle R_g \rangle$, nm | $\langle R_{ee} \rangle$, nm | $\langle R_h \rangle$, nm | $\langle E \rangle$ |
|---|---|---|---|---|
| A$\beta$40 | 1.65 ± 0.012 | 3.83 ± 0.065 | 1.47± 0.01 (1.6)* | 0.65 ± 0.014 (0.6)† |
| A$\beta$42 | 1.69 ± 0.013 | 3.93 ± 0.066 | 1.51 ± 0.01 (1.7)‡ | 0.63 ± 0.015 (0.6)† |

The errors are reported in terms of standard deviations estimated using block averaging.
*The experimental value of $R_h$ is from FCS experiments of Wennmalm et al. (54).
†The FRET efficiency values are from Meng et al. (30). The numbers in parentheses are taken from the quoted experiments
‡The experimental value of $R_h$ is from the diffusion NMR experiments of Vendruscolo and coworkers (53).

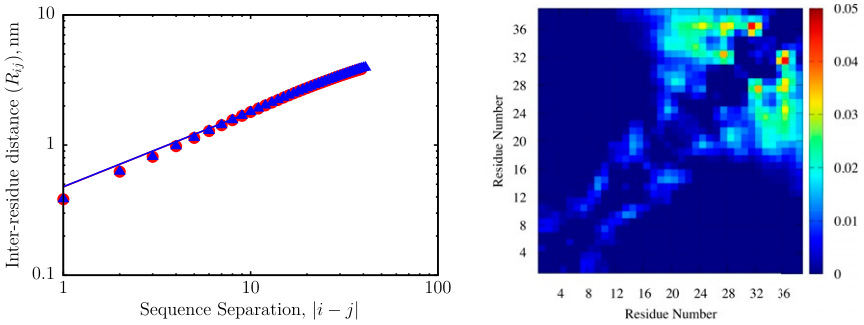


**Fig. 2.** (*Left*) The variation of interresidue distances ($R_{ij}$) with sequence separation, $|i-j|$ for the A$\beta$40 and A$\beta$42 sequences, are shown as red circles and blue triangles, respectively. The solid lines are fits to the Flory scaling law expected for homopolymers in good solvents: $R_{ij} = R_0|i-j|^{\nu}$, with $R_0$ = 0.47 nm and $\nu$ = 0.58 for both the peptides. Although $R_{ij}$ scaling, an ensemble-averaged quantity, is consistent with polymer theory, the distribution of $R_{ee}$ deviates from the expected universal behavior for self-avoiding walks (Fig. 1). (*Right*) Variation in the ensemble-averaged side-chain contacts for the A$\beta$40 and A$\beta$42 shown as a difference map, color coded (scale on the right) in terms of $\delta p_{ij} = \langle p_{ij}^{A\beta42}\rangle - \langle p_{ij}^{A\beta40}\rangle$. Here, $\langle p_{ij}^{A\beta40}\rangle$ and $\langle p_{ij}^{A\beta42}\rangle$ are the ensemble-averaged contact probability maps for A$\beta$40 and A$\beta$42, respectively.

consistent with previous work (18, 27, 48). It is worth emphasizing that the probabilities to adopt $\beta$-sheet or $\alpha$-helix are really small relative to coil-like states. Our simulations suggest that detecting them using experiments is likely to be difficult.

**Chemical Shifts.** The residue-wise chemical shifts for A$\beta$40 and A$\beta$42 are shown in Fig. 4. Given the coarse-grained nature of our model, and the higher errors associated with the estimation of chemical shifts of other nuclei with the LARMOR-$C_\alpha$ formalism (*SI Appendix*) (61), we report only the ensemble-averaged values of the $C_\alpha$ chemical shifts. As illustrated in Fig. 4, our predictions are in reasonable agreement with NMR experiments (17, 18). We note that the qualitative agreement was obtained (*SI Appendix*, Fig. S6) without adjusting any parameter to fit the experiments. To further validate the SOP-IDP model, we carried out a systematic comparison between our predictions for the $C_\alpha$ chemical shifts and those from all-atom simulations based on different force fields (48, 62). The relative errors with respect to experimental data show that the predictions of the SOP-IDP model are on par and in some cases better than the all-atom force fields (see *SI Appendix*, *SI Text*, Table S3, and Figs. S7 and S8 for details).

The chemical shifts for the first 35 residues are nearly identical for both the peptides (Fig. 4 and *SI Appendix*, Fig. S9), mirroring the observed trend for the secondary structure profiles. Most residues show only minimal departures from the RC references, further demonstrating the completely disordered nature of the ensembles and the lack of substantially populated (metastable) structural elements. Nonetheless, the chemical shifts for some valines (Vs) and histidines (Hs) deviate from the RC values. In particular, the upshifted $C_\alpha$ shifts for V18 (also visible in the experimental profiles) and V12-H14 for both A$\beta$ monomers imply that this region could adopt a residual $\beta$ structure, albeit only transiently. Similar observations have been made previously (63) using electron paramagnetic resonance (EPR) spectroscopy and all-atom simulations (34, 48). The chemical shifts for H6, V24, and V36 are also shifted upfield by $\approx$1 ppm, indicating that these residues could form fleeting $\beta$ structures. Although similar signatures are not present in the experimental profiles shown in Fig. 4, a previous study (48) does suggest that these residues could be involved in hydrogen-bonding interactions with V18, resulting in upshifted chemical shifts. Interestingly, the modest departure from the RC values for H6 and H14 hints at a possible role of these residues in mediating A$\beta$ self-assembly (18, 64) via formation of transient $\beta$ structures, as well as being a template for molecular recognition (65).

Given that all of the measurable and computable values are virtually identical for both the peptides, is it possible to explain the enhanced aggregation rate of A$\beta$42 relative to A$\beta$40 using equilibrium monomer properties? We show below that the answer lies in the differences in the conformational heterogeneities between the two peptides, which result in an enhanced population of $N^*$ states in A$\beta$42 compared to A$\beta$40.

**Sequence-Dependent Conformational Heterogeneities.** The ensemble-averaged properties, as well as features of the free energy landscapes, although commensurate with the prevailing consensus regarding the intrinsically disordered nature of the A$\beta$ monomeric ensembles, do not reveal significant alloform-specific differences. Structural details of aggregation-prone conformers having extremely low equilibrium weights are masked in such averages, although subtle differences in the secondary structure profiles allude to sequence-specific conformational heterogeneity. To glean further insights into the fine structure of the free energy landscape of A$\beta$40 and A$\beta$42, we carried out hierarchical clustering of the conformational ensembles based on a robust distance metric (*SI Appendix*, Fig. S10). Our previous study (52) showed that this scheme provides an efficient means to quantify the contrasting conformational heterogeneities of many IDPs.

At a coarse level, the conformational ensembles of both the peptides can be partitioned into three major clusters: compact, semicompact, and extended. Even when such a simplistic classification is used, sequence-specific conformational preferences become apparent (*SI Appendix*, Fig. S10). For both the sequences, the equilibrium weights of the fully extended structures are relatively low, being 19.9 and 20.9% for the A$\beta$40 and A$\beta$42 sequences, respectively. Compact structures dominate the conformational ensemble of A$\beta$40 (53.7%), and semicompact structures constitute the second-largest cluster (26.4%). This trend is reversed for A$\beta$42; the population of semicompact structures is enhanced to 41%, and that of compact structures is diminished to 38.1%. The relative weights of the different cluster families, even at such a coarse resolution, hint at evident deviations from standard polymer models, even though ensemble-averaged contact maps and scaling behavior suggest otherwise.

To ascertain the arrangement of the local structural segments we performed more refined analyses using a smaller distance cutoff for $D_{ij}$ (see Eq. **11** in *SI Appendix*). This splits the dendrogram further and divides the major clusters into additional subfamilies. A cutoff of 5.0 seems appropriate for both the A$\beta$40 and A$\beta$42 ensembles as it not only provides sufficient resolution to determine the key differences in the contact maps of the

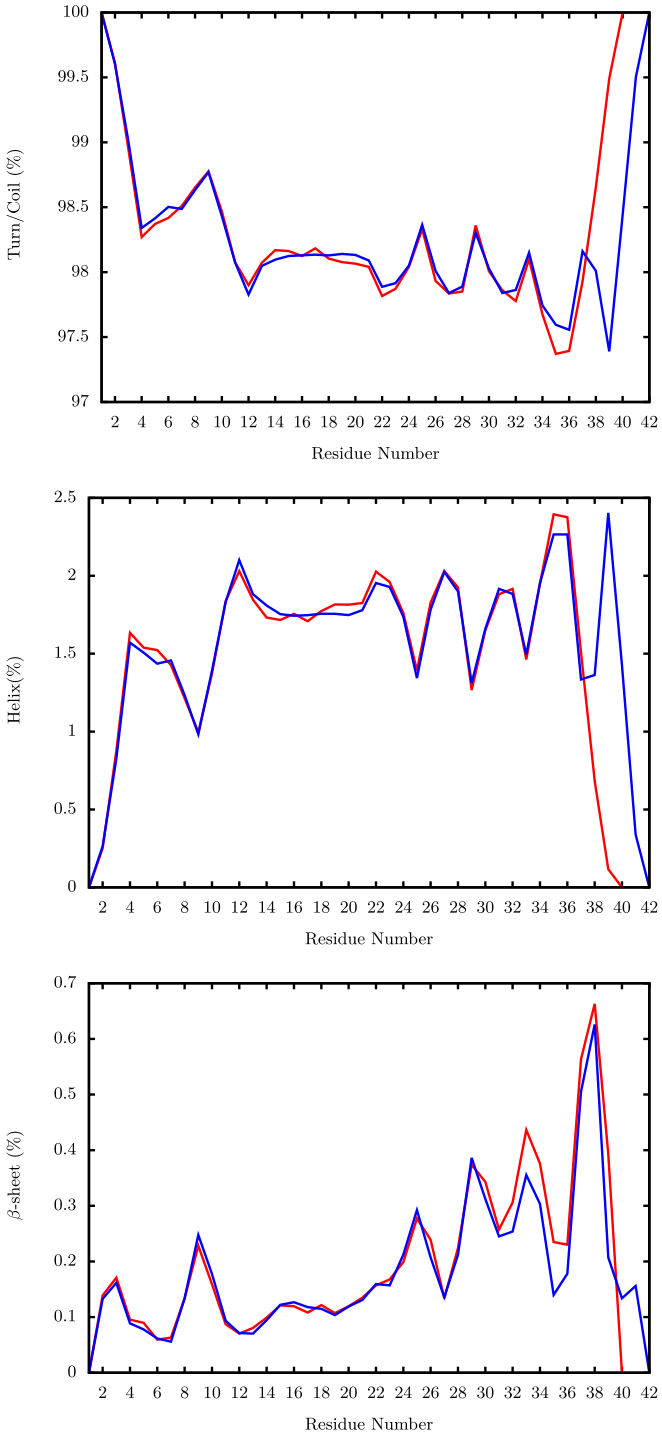


**Fig. 3.** Percentages of different secondary structure elements in the conformational ensembles of A$\beta$40 (red) and A$\beta$42 (blue). (*Top*) Ensemble-averaged percentages for forming turn/irregular coils. (*Middle*) Residue-wise percentages for forming helices. (*Bottom*) Residue-wise percentages for forming $\beta$-sheets. The secondary structures were assigned based on the positions of the $C_\alpha$ atoms using the protein C-alpha secondary structure output (PCASSO) algorithm.

subclusters, but also keeps their number tractable. Using this scheme, we obtained 12 clusters for A$\beta$40 and 11 clusters for A$\beta$42 (*SI Appendix*, Fig. S10). The populations of the different clusters are tabulated in *SI Appendix*, Table S4. In Fig. 5 we show the most important cluster families that are related to the $N^*$ states, and the rest are included in *SI Appendix*, Figs. S12 and S13. The proper segregation of the various families in the two-dimensional surfaces defined by $R_g$ and $R_{ee}$ implies that our clustering scheme is robust (*SI Appendix*, Fig. S11). The structural details of the different subfamilies are described below.

***A$\beta$40.*** The contact maps associated with clusters 1, 2, 3, and 7 (*SI Appendix*, Fig. S12) are devoid of short- or long-range contacts. Other clusters are also rich in RC-like conformations, but also contain some fingerprints in the contact map, which is suggestive of local structuring near the N terminus. Approximately 50% of the conformations in clusters 5 and 8 show enhanced structural contacts between the N terminus (residues ≈1 to 8) and those near the CHC (*SI Appendix*, Table S4). Specifically, residues ≈12 to 21 in cluster 5 and ≈10 to 20 in cluster 8 engage in interactions with the N terminus. The contact maps of clusters 4 and 6 (*SI Appendix*, Fig. S12) also indicate enhanced ordering near the N terminus of A$\beta$40. However, these features are less prominent than those in clusters 5 and 8, and the contacts tend to be more short ranged (between residues $i$ and $i+3$). Cluster 12 consists of a nearly equal population of RC-like conformations and those that have a number of long-range contacts between the N terminus and residues ≈15 to 28 (*SI Appendix*, Fig. S12 and Table S4). Both simulations (22, 23, 66) and experiments (59, 60, 67, 68) have previously underscored the importance of residual structure in this region (particularly, LVFFA, residues 17 to 21) and its potential role in dictating aggregation kinetics. Overall, clusters containing exclusively RC-like states account for 32% of the equilibrium population. The cumulative population of clusters having a mixture of RC-like conformations and those having residual structuring near the N terminus is 47.3%. Taken together, ≈80% of the conformational ensemble is predominantly RC-like.

***Fingerprints of the fibril structure.*** Despite the dominance of RC-like conformations in the equilibrium ensemble, the essential fingerprints of the fibril state can be found in a few conformational clusters. Importantly, at the distance cutoff used for structural clustering, none of the constituent structures in these clusters resemble random coils. Previous studies have shown that the D23-K28 salt bridge, which is a key structural element of mature A$\beta$40 fibrils, modulates the early events of amyloid aggregation (20, 21, 31). We find clear-cut signatures of these contacts in the structures in cluster 9 (Fig. 5*B*), which is associated with an equilibrium population of 7.7%. In addition to the presence of E22/D23-K28 salt bridges, a majority of structures also have hydrophobic contacts between V24 and N27, which stabilize the VGSN turn. However, the N-terminal and C-terminal segments remain mobile and do not form the complete strand–loop–strand (SLS) structure, which forms the repeating unit in the fibril. The conformations in cluster 11 (population 10%) are compact (*SI Appendix*, Fig. S11) with prominent off-diagonal elements in the contact map, which are reflected in the tendency to form SLS-like structures (see *SI Appendix* for details) via association of the two ends (Fig. 5*D*). The SLS structure, often thought to be a putative intermediate in fibril elongation (69), essentially has a hairpin-like topology and engages in hydrogen-bonding interactions with neighboring monomers to stabilize the fibril structure. Many of the conformations within cluster 10 (population 3.2%) are ordered and exhibit signatures of the fibril state (Fig. 5*C*). In these structures, contacts are formed between the CHC and residues 28 to 40.

***A$\beta$42.*** Conformations belonging to clusters 2, 3, 8, and 9 consist of exclusively RC-like structures. Some clusters identified for A$\beta$42 consist of an admixture of RC-like conformations and ones with different extent of local structuring (*SI Appendix*, Table S4). In cluster 1, nearly two-thirds of the constituent structures feature contacts between the central region (residues ≈22 to 29) and the C terminus (*SI Appendix*, Fig. S13). Structures within

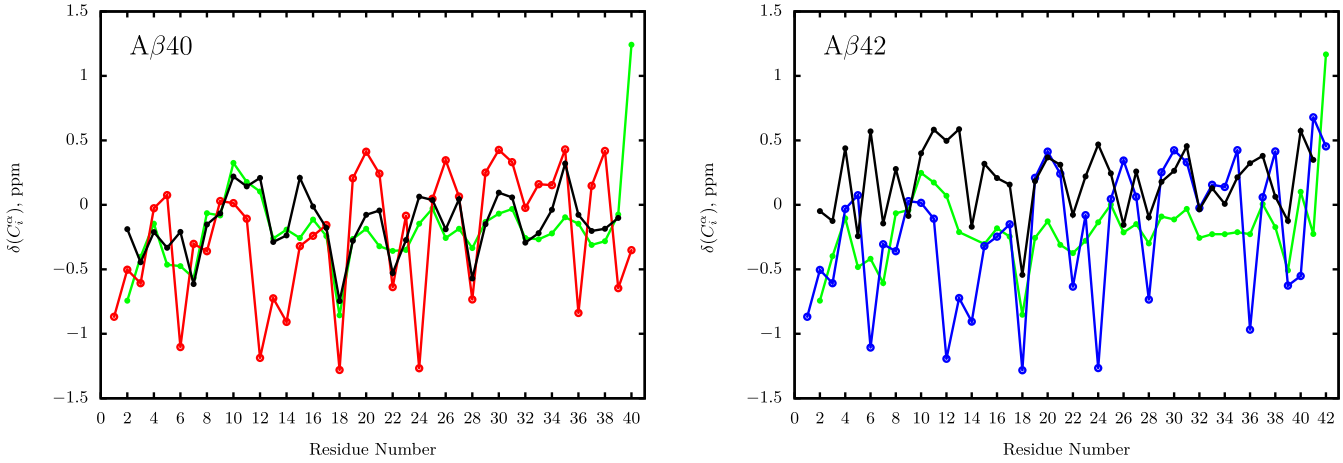


**Fig. 4.** Residue-wise $C_\alpha$ chemical shifts for A$\beta$40 (*Left*) and A$\beta$42 (*Right*). The predicted shifts for A$\beta$40 are shown in red, and those for A$\beta$42 are shown in blue. The chemical shifts from the experiment of Bax and coworkers (17) are shown in light green, and those from Zagorski and coworkers (18) are shown in black. The predicted values are in qualitative (the trends are comparable) agreement with experiment. The weighted-root-mean-square error ($\zeta$) for A$\beta$40 between predicted shifts and those measured by Bax and coworkers (17) is 0.53. The error is smaller (0.47) relative to the measured values of Zagorski and coworkers (18). For A$\beta$42, the trend is reversed, and corresponding $\zeta$ values are 0.48 and 0.66.

cluster 10 are mostly semicompact, and around 20% of these exhibit some tendency of structuring near the N terminus by forming contacts with residues ≈20 to 30 (*SI Appendix*, Fig. S13 and Table S4). Conformations within cluster 11 are mostly disordered and are typically RC-like, although a few of them exhibit structuring near the central region (*SI Appendix*, Fig. S13). Together, clusters having predominantly RC-like features (albeit with some degree of local structuring) contribute around 62.1% to the equilibrium population (Fig. 5*A*). In addition to those having RC-like features, there are other clusters in the conformational ensemble of A$\beta$42, which exhibit structural features of different experimentally characterized fibril morphologies. We describe them below.

***Fingerprints of the S-bend fibril structure.*** Recently, NMR (44, 46) and cryo-EM (70) experiments have identified the "S-bend" motif as a building block of the A$\beta$42 fibril structure. The S-bend structure, which apparently is not found in A$\beta$40, results from enhanced ordering near the C terminus of A$\beta$42 and is characterized by a close contact between residues K28 and A42. Although not immediately discernible from the contact maps, many of the conformations within clusters 4 and 5 exhibit the characteristic S-bend motif, stabilized by a close contact between residues K28 and A42 (Fig. 5). These clusters exhibit enhanced ordering near the C terminus, but differ somewhat in the specific residues involved in contact formation. In cluster 4 (Fig. 5*E*) the contacts are formed between the C terminus and residues 18 to 24, while in cluster 5 (Fig. 5*F*) residues 15 to 22 form contacts with the C terminus. Conformations similar to these have been previously identified by Garcia and coworkers (56) using all-atom simulations in conjunction with spectral clustering algorithms.

***Fingerprints of the U-bend fibril structure.*** Besides the S-bend topology, A$\beta$42 also forms fibril structures in which the peptide adopts the canonical U-bend (71) (similar to the repeating unit of A$\beta$40 fibrils). Clusters 6 (Fig. 5*G*) and 7 (Fig. 5*H*) exhibit SLS-like structures (see *SI Appendix* for details) and consist of the E22/D23-K28 salt bridges, as well as the VGSN turn in a large fraction of the constituent structures. These clusters contribute 9.7 and 10.2% to the equilibrium population. In cluster 6, the $\beta$-strand–loop–$\beta$-strand structures are formed via the association between residues ≈9 to 16 and the C terminus, whereas SLS conformations within cluster 7 involve contacts between the two terminii.

From our clustering analyses we draw important insights regarding the sequence-specific heterogeneities of A$\beta$40 and A$\beta$42 monomers: 1) For both the sequences, clusters that are predominantly RC-like or have RC-like structures in an admixture with conformations having some degree of residual structuring dominate the equilibrium population. Notably, these clusters do not exhibit any of the characteristic structural features found in the fibril. 2) Interestingly, despite having identical sequence for the first 40 residues, the structural features near the termini are distinct for the two A$\beta$ monomers. A$\beta$40 has a structured N terminus, while A$\beta$42 exhibits enhanced structuring near the C terminus, which is consistent with previous studies (19, 42, 48, 72, 73). As argued in a previous work (42), in A$\beta$ monomers, there is a tug of war between the C terminus and the N terminus for making contacts with the CHC. In the case of A$\beta$40, the N terminus wins, because unlike A$\beta$42, it does not contain an extra hydrophobic patch formed by I41 and A42 to induce structuring near the C terminus. Structuring near the N terminus of A$\beta$40 could also have implications in the subsequent assembly process and in determining the fibril morphology. As shown in a recent study (74), a cryo-EM A$\beta$40 fibril structure derived from a meningeal Alzheimer's tissue is characterized by pronounced arches at the two ends (most notably at the N terminus). 3) Although masked in ensemble averages, clustering analyses reveal that the aggregation-prone structures ($N^*$ states) are present in the spectrum of conformations populating the ensemble, in agreement with previous studies (21, 32, 34, 36). For A$\beta$40, we find evidence of formation of E22/D23-K28 salt bridges, as well as SLS topologies, and within the A$\beta$42 ensemble, we find signatures of SLS structures characterized both by U-bends and by the S-bend motifs.

**Linking the Population of $N^*$ States to Aggregation Propensity.** The analyses based on hierarchical clustering show that fibril-like or aggregation-prone conformations ($N^*$) are populated in the monomer ensemble of the A$\beta$ peptides. However, to quantify their population, $p_{N*}$, a stringent order parameter, comparing the putative $N^*$ conformations identified using the cluster analysis with the fibril structure, is required. In other words, only a subset of conformations belonging to the $N^*$ basin of attraction has the propensity to aggregate. We estimate $p_{N*}$ based on a geometric criterion. Because the conformations belonging to the $N^*$ state should be similar to the structure in the fibril, we use the

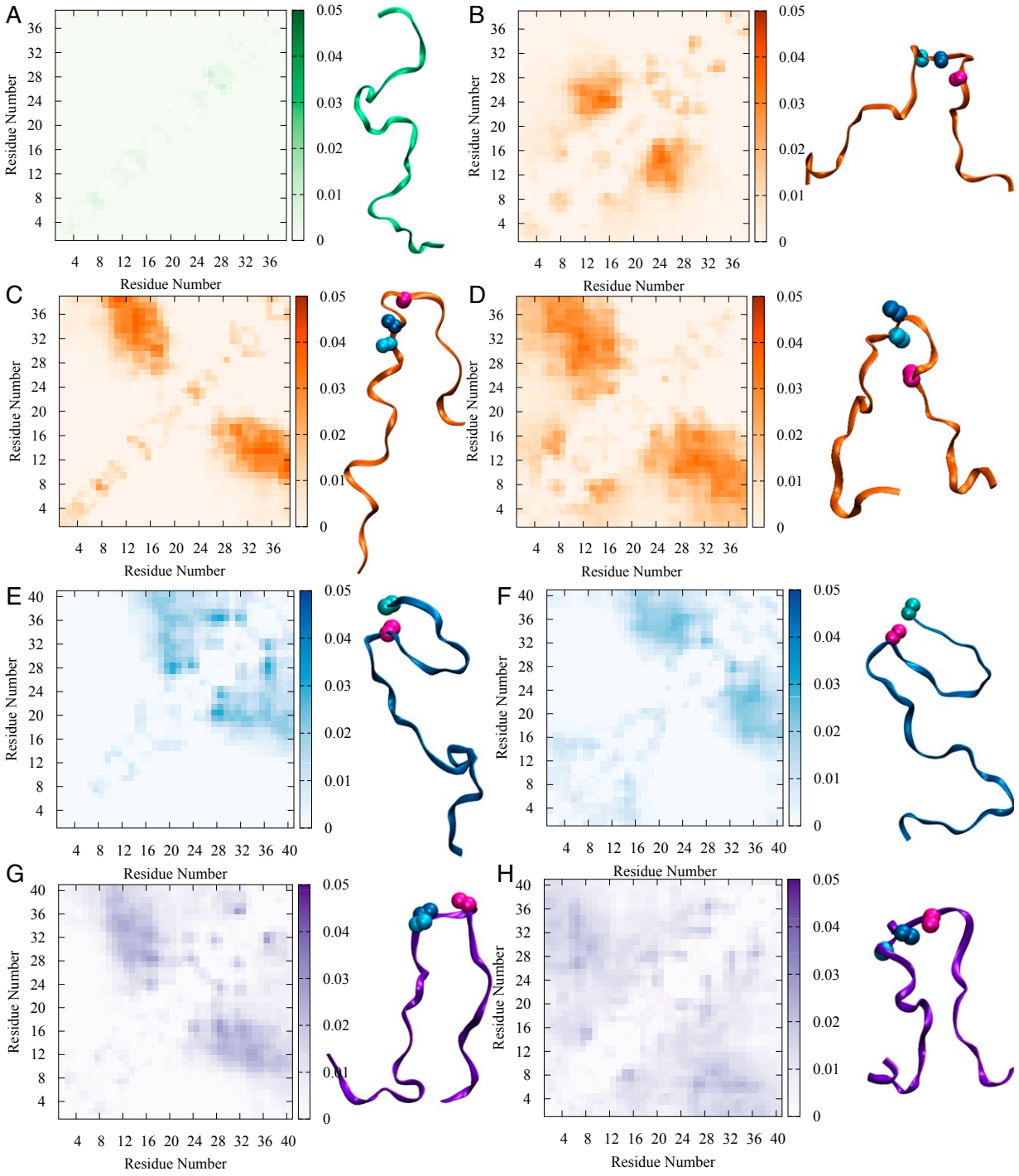


**Fig. 5.** (*A*) The free energy ground state of the A$\beta$40 and A$\beta$42 ensembles consisting of RC-like conformations devoid of fibril-like contacts. The contact map averaged over the corresponding subensemble is featureless. *B–D* show the contact maps and representative snapshots of clusters identified for A$\beta$40, which consist of aggregation-prone ($N^*$) conformations. These conformations exhibit many of the structural elements found in the A$\beta$40 fibril. The E22, D23, and K28 residues, which form the key salt bridges, are explicitly shown. (*B*) Cluster 9: The VGSN turn is formed and is stabilized by the D23-K28 salt bridge. However, the N-terminal and C-terminal segments are quite flexible and do not form contacts. (*C*) Cluster 10: A SLS-like structure stabilized by the formation of a VGSN turn and the D23-K28 salt bridge. (*D*) Cluster 11: SLS-like structure stabilized by a turn formed near the VGSN region and the E22-K28 salt bridge. *E–H* show the contact maps and representative snapshots of clusters identified for A$\beta$42, which consist of conformations, in some of which the structural elements found in the different A$\beta$42 fibril polymorphs are visible. These structures are members of the $N^*$ ensemble. In *E* and *F* the K28 and A42 residues are shown as spheres. (*E*) Cluster 4: S-bend type motif, exhibiting close contact between K28 and A42, similar to that found in an A$\beta$42 fibril structure. (*F*) Cluster 5: S-bend type structures, with enhanced ordering near the C terminus, including a close contact between the K28 and A42 residues. *G* and *H* denote clusters which bear resemblance to the monomer unit of a different A$\beta$42 polymorph, which has a classical $\beta$-strand–loop–$\beta$-strand topology consisting of a U-bend. (*G*) Cluster 6: SLS-like structure consisting of a U-bend near the VGSN turn region, stabilized by the D23-K28 salt bridge (shown as spheres). (*H*) Cluster 7: SLS-like structures formed via the association of the two terminii. The D23-K28 salt bridge (shown as spheres) stabilizes the turn region in these structures.

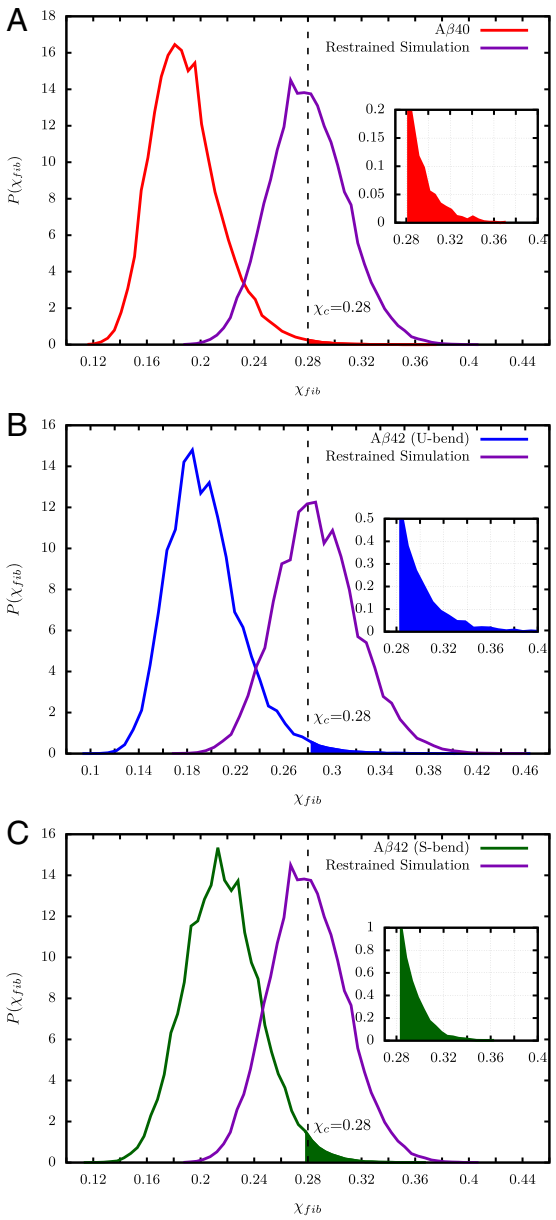


**Fig. 6.** (*A*) Distribution of $\chi_{fib}$ for A$\beta$40 (shown in red). The overlap parameter is computed with respect to the monomers in the experimental fibril structure. For A$\beta$40, the reference state is the monomer unit from the striated (U-bend) fibril structure (75) (Protein Data Bank [PDB] ID: 2M4J) (75). The shaded area under the curve, which corresponds to the $N^*$ state is in red. *Inset* shows an enlarged view of this section of the curve. (*B*) Distribution of $\chi_{fib}$ for A$\beta$42 (shown in blue) using the monomer from the U-bend fibril structure (PDB ID: 2BEG) (71) as a reference. The shaded area under the curve (*Inset*) corresponds to the population of $N^*$ states with U-bend topology. (*C*) Same as *B* except that the overlap parameter, $\chi_{fib}$, is calculated using the S-bend fibril structure (PDB ID: 2NAO) (44). The shaded area under the curve (*Inset*) corresponds to the population of $N^*$ states having S-bend topology. In *A*–*C* the purple curves correspond to $P(\chi_{fib})$ computed from the restrained simulations (see *SI Appendix* for details) starting from the monomer present in the fibril structure. (*A*) A$\beta$40 fibril structure (PDB ID: 2M4J). (*B*) A$\beta$42 U-bend fibril structure (PDB ID: 2BEG). (*C*) A$\beta$42 S-bend fibril structure (PDB ID: 2NAO). In all cases, the center of the distribution is approximately at $\chi_{fib}$ = 0.28. This value corresponds to $\chi_c$ and is marked in the plots.

fibril structures as reference states. We used the structural overlap parameter, $\chi_{fib}(t)$, between a monomer conformation generated in the simulations at time $t$ and the monomer in the experimental fibril structure to characterize the $N^*$ states for both the peptides (see *SI Appendix* for further details). For A$\beta$40, we used the striated fibril structure determined by Tycko and coworkers (75) as the reference. For A$\beta$42, we used the two structures reported by Riek and coworkers: the U-bend topology (71), which is very much similar to the repeating unit found in A$\beta$40 fibrils, and the S-bend motif (44).

The distributions of $\chi_{fib}$ are shown in Fig. 6. The distributions for A$\beta$40 and A$\beta$42 nearly overlap when the U-bend (or striated) structure is chosen as a reference. This suggests that the population of fibril-prone structures exhibiting the canonical SLS motif is approximately similar between the two sequences. In contrast, the conformational ensemble of A$\beta$42 seems to bear a higher overall resemblance to the S-bend motif.

The $N^*$ state is hidden in the tail of these distributions (Fig. 6). To identify the approximate location of the basin consisting of the aggregation-prone conformations, we carried out restrained simulations (see *SI Appendix* for details) starting from the monomer structure present in the experimental fibril morphologies of A$\beta$40 and A$\beta$42. The $\chi^R_{fib}$ distributions are calculated from the restrained simulations and reflect the conformations within the $N^*$ basins of attraction that could readily aggregate (purple curves in Fig. 6 *A*–*C*). We find that the basin of aggregation-prone structures is approximately centered at $\chi_c$ = 0.28 for both A$\beta$40 and A$\beta$42. Hence, for both the sequences, we collectively define $N^*$ states as the ensemble of conformations with $\chi_{fib} \geq \chi_c$. Therefore, we calculated $p_{N^*}$ using

$$p_{N^*} = \frac{1}{N_j}\sum_{j=1}^{N_j}\frac{1}{N_s}\sum_{s=1}^{N_s} H\left(\chi_s^j - \chi_c\right)\Big), \qquad [1]$$

where $\chi_s^j$ is the structural overlap for the $s$th snapshot in the $j$th trajectory, $N_s$ denotes the number of snapshots in each trajectory, $N_j$ denotes the number of independent trajectories, and $H$ is the Heaviside function. Using Eq. **1**, we estimated that $p_{N^*}$ is 0.45% for A$\beta$40. For A$\beta$42, $p_{N^*}$ is estimated to be 1.45% when the U-bend topology is used as a reference and 2.17% when the S-bend motif is used as a reference. These numbers fall within the same range as previous estimates of $p_{N^*}$ at physiological conditions for the src SH3 (33) domain, which is a globular protein.

As indicated by the presence of a minor peak near $\chi_c$, only a subpopulation of conformations within the clusters (*SI Appendix*, Fig. S17), which exhibit structural signatures of the fibril state, meets the strict geometric definition (Eq. **1**) for the conformations within the $N^*$ state that are likely to aggregate. This observation implies that although contact maps corresponding to various subensembles could be useful in discerning the presence of aggregation-prone species, they are probably too coarse as metrics for quantitatively measuring $p_{N^*}$.

Previously (32), we proposed that the time scale of fibril formation, which should be thought of as an approximate estimate of the mean first passage time for the monomer to reach the structure in the fibril (U-bend or S-bend), is related to $p_{N^*}$ as

$$\tau_{fib} \propto exp(-Cp_{N^*}), \qquad [2]$$

where the population of the $N^*$ state is expressed as a percentage. The expression in Eq. **2** does not account for the multiplicity of events that are involved in fibril formation. Using $C$ = 1 for both the peptides and the estimated values of $p_{N^*}$, we predict that $\tau_{fib}$ for A$\beta$42 is approximately 24 times smaller than for A$\beta$40. Our prediction is in accord with recent experiments (39), which show that the fibril formation rate of A$\beta$42 is about one order of magnitude greater than that of A$\beta$40. Interestingly, the one order of magnitude difference cannot be rationalized if only

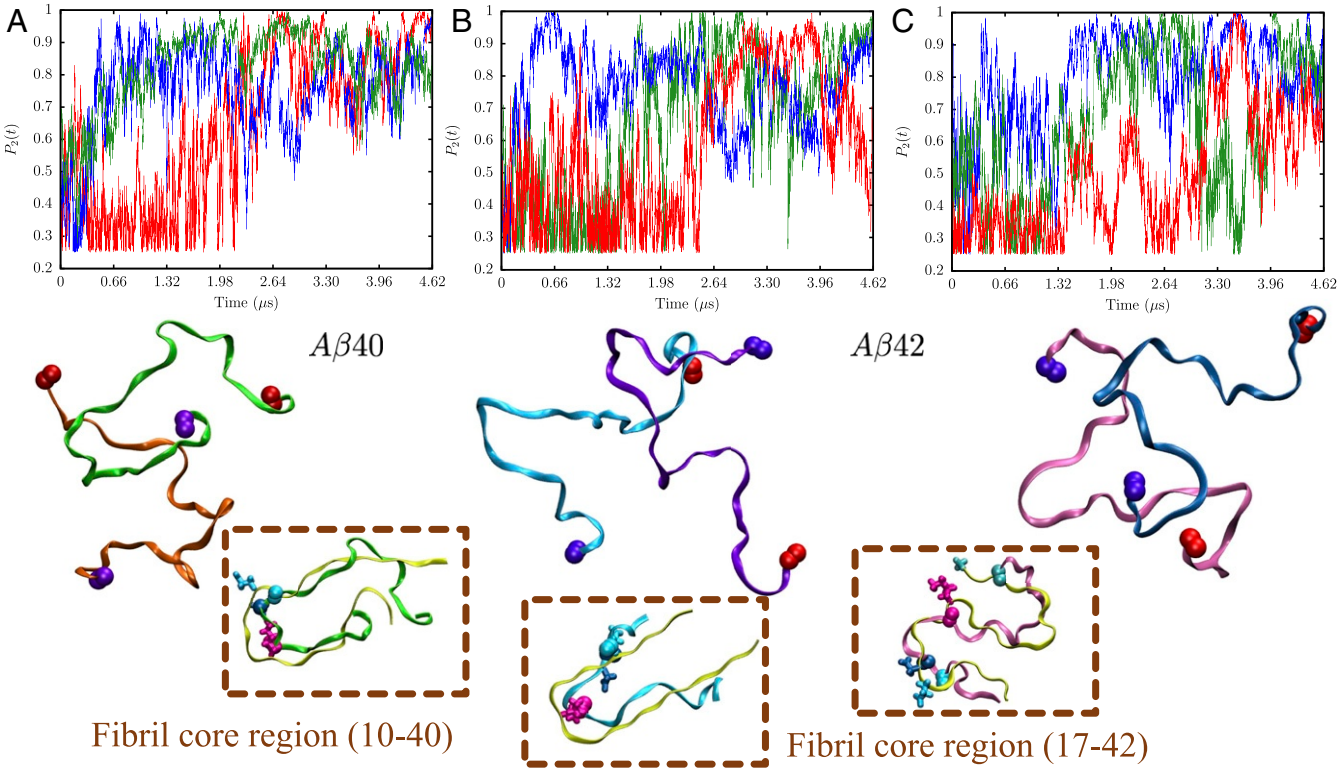


**Fig. 7.** Kinetics of dimerization observed for A$\beta$40 and A$\beta$42 when two $N^*$ states are allowed to coalesce. (*A*) A$\beta$40. *Top* shows the time evolution of the nematic order parameter, $P_2$ (high $P_2$ implies fibrillar order). Three representative trajectories associated with different first passage times of dimer formation are shown. *Bottom* shows a representative snapshot of the A$\beta$40 dimer. (*B*) *Top* shows the time evolution of $P_2$ for three trajectories initiated from two $N^*$ conformations of A$\beta$42 having the U-bend topology. *Bottom* shows a representative snapshot of the corresponding dimer. (*C*) *Top* shows the time evolution of $P_2$ for three trajectories initiated from two $N^*$ conformations of A$\beta$42 having the S-bend topology. *Bottom* shows a representative snapshot of the corresponding dimer. In all of the snapshots, the red spheres denote the N terminus and the violet spheres denote the C terminus of the two chains. For all of the dimer snapshots shown in *A*–*C*, *Insets*, a zoomed-in view of the fibril core region is also shown. For both A$\beta$40 and A$\beta$42, one of the chains in the dimer overlaps appreciably with the monomer unit of the fibril structure (shown in yellow), while the other chain fluctuates to maintain fluidity. The E22, D23, and K28 residues which participate in salt bridge formation are shown in cyan, blue, and magenta colors, respectively. In case of the S-bend motif of A$\beta$42, residue A42 is also shown.

the U-bend or striated conformations are identified with the $N^*$ state. In this case, the aggregation of A$\beta$42 is only twice as fast. This implies that the enhanced ordering near the C terminus, which stabilizes the S-bend motif, provides an alternate route for fibril formation in A$\beta$42 and could possibly explain why it is more aggregation prone despite being present in a much lower concentration than A$\beta$40 in the plasma of cells.

***N**** **States Spontaneously Assemble into Dimers.** To determine whether the monomer excited states are harbingers of aggregation we simulated the dynamics two $N^*$ states oriented randomly with respect to one another. The nematic order parameter, $P_2$, was used to quantify dimer formation. For both A$\beta$40 and A$\beta$42, two $N^*$ conformations readily coalesce to form dimers upon collision, on the time scale of microseconds (Fig. 7). Melting and reformation of some of the dimer conformations hint at fluid-like behavior, which suggests that the critical nucleus is larger than two.

We also carried out simulations where two $N^*$ states were aligned nearly perfectly relative to one another. These form dimers rapidly and mostly retain their fibril-like order. Some configurations exhibit only minor orientational melting (*SI Appendix*, Fig. S18). These exploratory simulations provide a sneak peek into the diverse range of dimer conformations explored by the A$\beta$ peptides. For both the peptides, the dimer interface is heterogeneous (*SI Appendix*, Fig. S19), involving interactions between different segments of the monomer chains. Specifically, for A$\beta$42, the C terminus seems to play a key role in mediating monomer–monomer interactions.

When the simulations were initiated from monomer conformations, which do not exhibit fibril-like characteristics (non-$N^*$), no appreciable fibril-like order was found on the time scale of microseconds (*SI Appendix*, Fig. S18). In a few trajectories, we observed only transitions to a state with intermediate order (characterized by $P_2 \approx 0.6$). Therefore, the free energy barrier for dimerization is likely to be much higher for these states.

## Discussion

We used the SOP-IDP model to characterize the conformational ensembles of A$\beta$40 and A$\beta$42 monomers to quantify the role their excited states play in controlling fibril formation. The ensemble averages of many properties are practically identical for the two sequences. Clustering analyses revealed the details of sequence-specific heterogeneity. The interresidue contact maps for individual clusters show that both the peptides access an array of different structures. Specifically, the A$\beta$40 monomer has a measurable probability of being structured at the N terminus, while the C terminus of A$\beta$42 is ordered to some extent. For both the sequences, we find prominent fingerprints of the existence of aggregation-prone (excited) conformations ($N^*$ states), which are only sparsely populated. The $N^*$ states share a number of features in common with the monomer structures in the fibrils (Fig. 8). By using the differences in the population of $N^*$ states between the two peptides in Eq. **2**, we rationalize the higher value

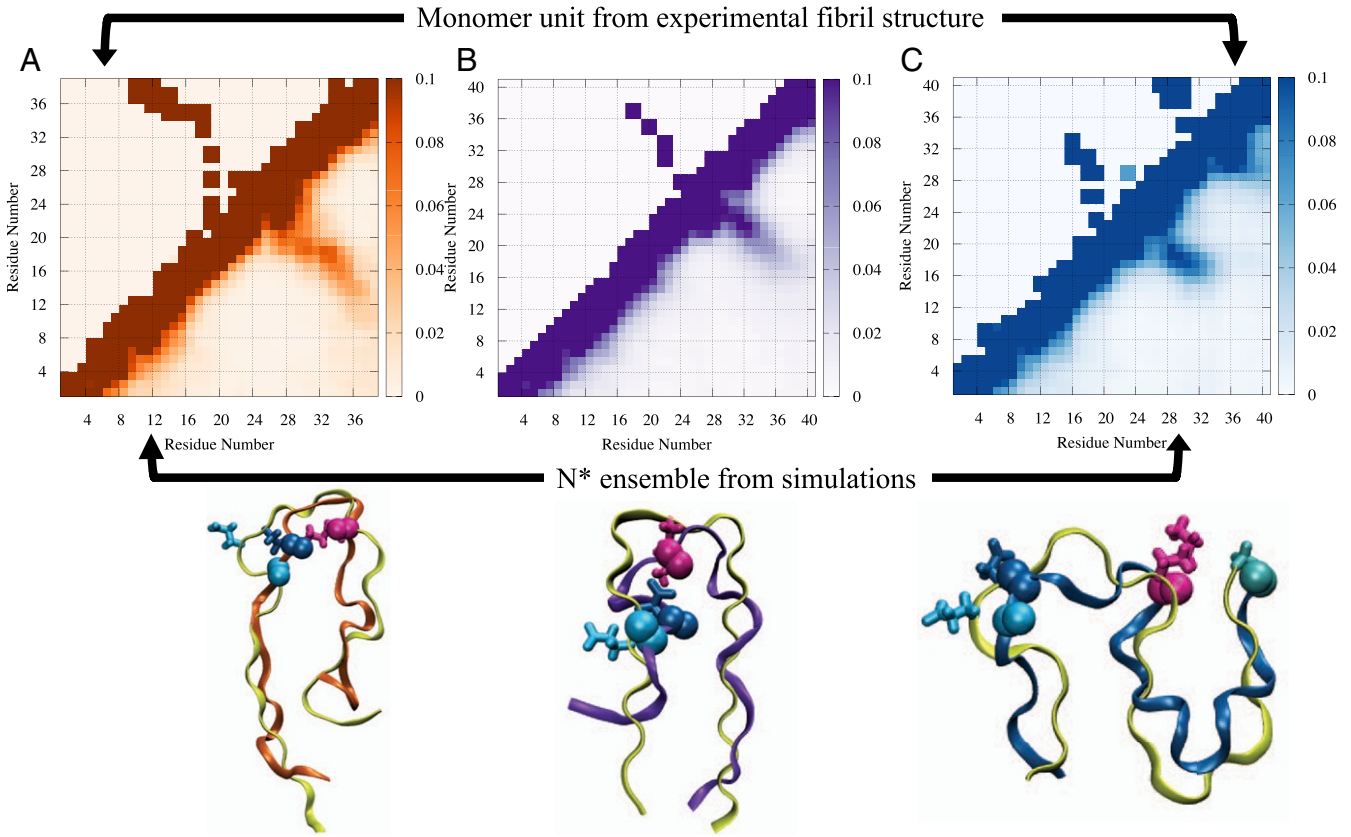


**Fig. 8.** (*A–C, Top*) Contact maps for the $N^*$ ensemble identified from simulations (lower triangle) and the monomer unit of the experimental fibril structure (upper triangle). (*A*) $N^*$ ensemble of A$\beta$40, consisting of SLS-like configurations, stabilized by the VGSN turn, and D23-K28 salt bridge. As is evident, these structural features are also found in the A$\beta$40 fibril structure. At least two structurally distinct $N^*$ states exist for A$\beta$42, as shown in *B* and *C*. (*B*) SLS-like structure consisting of a U-bend near the VGSN region, stabilized by the D23-K28 salt bridge. The positions of the D23-K28 salt bridge in the simulated $N^*$ ensembles overlap to a great extent with the experimental structures in *A* and *B*. (*C*) S-bend motif stabilized by a contact between K28 and A42. (*A–C, Bottom*) Below the contact maps, a representative structure from the $N^*$ ensemble is shown superimposed on the monomer unit from the fibril structure (shown in yellow). The E22, D23, and K28 residues are shown in cyan, blue, and magenta colors, respectively. For the S-bend motif, residue A42 is also shown. For clarity, we only show residues that are part of the fibril core (10 to 40 for A$\beta$40 and 17 to 42 for A$\beta$42).

of $\tau_{fib}$ in A$\beta$40 relative to A$\beta$42. We find that two $N^*$ states coalesce rapidly to form dimers. Thus, $N^*$ states may be templates for self-assembly.

The propensity to form $\beta$ structures in the spectrum of monomers could be used as a marker for protein aggregation. Indeed, the slightly upshifted values of the chemical shifts for some residues relative to the ensemble averages (*SI Appendix*, Fig. S20) do imply that the overall $\beta$ propensity is somewhat enhanced in the $N^*$ ensemble. Nonetheless, the $N^*$ state retains some degree of disorder, which is essential for the rapid switching between the different structures. Ordered $\beta$-sheet formation is expected only at the later stages in the aggregation cascade (formation of critical nuclei or protofibrils).

**Free Energy Gap between $N^*$ and Ground States Determines Aggregation Propensities.** The $N^*$ concept has important implications: 1) Because the $N^*$ state is separated from the ground state by a free energy gap, $\Delta G_{N^*}$, it follows that $p_{N^*}$ is vanishingly small if $\Delta G_{N^*}/k_B T >> 1$. In this scenario, aggregation is unlikely. 2) The value of $p_{N*}$ is slave to external conditions because $\Delta G_{N^*}$ can be modulated by varying the temperature, pH, or crowder concentration (14). This implies that the full sequence-dependent folding landscape of the monomer has to be determined to estimate $p_{N*}$. 3) Many proteins aggregate under conditions such that the $\Delta G_{N^*}/k_B T >> 1$ is not too large, which tidily explains why even the helical protein myoglobin can form fibrils under certain conditions (76). Sequences associated with higher values of $p_{N*}$ are therefore likely to have higher rates of aggregation. For instance, it is known that introduction of a lactam bridge near the VGSN turn region of A$\beta$ enhances the kinetics of fibril formation in A$\beta$40 by a factor of 1,000 (77), which was explained by substantial increase in the population of aggregation-prone $N^*$ states (21).

**Generality of the $N^*$ Theory.** If the lowest free energy state is ordered, aggregation to a fibril is possible only if the $N^*$ state is populated. This was illustrated by analyzing the experiments on the aggregation of the SH3 domain, a small globular protein. By using relaxation dispersion NMR experiments (78), it was shown that under ambient conditions a very low percentage of excited states are populated. Using simulations we showed that $p_{N*} \approx 2\%$, which allowed us to estimate the fibril elongation rate nearly quantitatively (33) using Eq. **2**. Significantly, the $N^*$ state determined in simulations was structurally identical to the one determined using NMR (see figure S7 in ref. 33 in which the two structures are superimposed). The earlier study (33) affirmed that even for globular proteins $p_{N*}$ provides a good estimate of fibril formation times and that the aggregation-prone structures have similarities to the monomer in the fibril.

This study expands the scope of the $N^*$ concept, from synthetic polymeric sequences and globular proteins to two assembly-competent biological sequences, namely A$\beta$40 and A$\beta$42, which

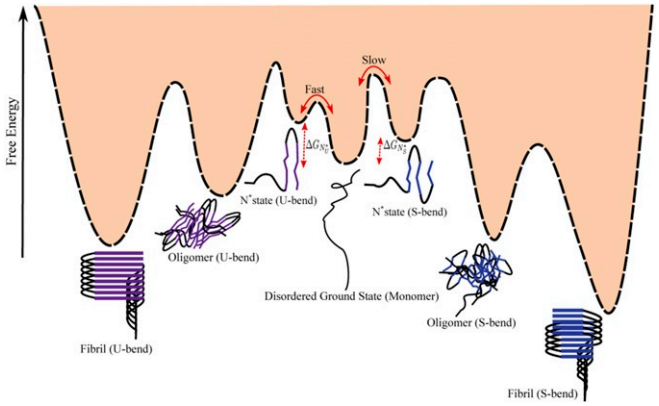


**Fig. 9.** A schematic free energy landscape illustrating the aggregation cascade that leads to the formation of fibrils, starting with the transition from the disordered ground state to the $N^*$ ensemble at the monomer level and ultimately the formation of fibril structures from an oligomeric assembly of $N^*$ states via different growth mechanisms. For A$\beta$42, the seeds of polymorphism are encoded in the structurally diverse $N^*$ states. Furthermore, the aggregation cascade seems to be in harmony with Ostwald's rule. For A$\beta$42, this would imply that $\Delta G_{N_U^*} > \Delta G_{N_S^*}$, and hence transition to the U-bend conformation is likely to be faster.

are inherently disordered in nature. The results show that the $N^*$ theory provides a molecular basis for the contrasting aggregation behavior of A$\beta$40 and A$\beta$42. Thus, the $N^*$ concept might be general and could provide important cues to the quantitative understanding of other aggregation-prone IDPs.

**Polymorphism and Ostwald's Rule.** The multiplicity of $N^*$ states (Fig. 8) implies that there should be several polymorphic fibril structures whose fingerprints should also be encoded in the excitation spectra of the monomer. The concepts underlying fibril polymorphism are illustrated using a schematic free energy landscape (Fig. 9). It is likely that the formation of distinct polymorphs follows Ostwald's rule. A recent study demonstrated how Ostwald's rule, well known in the formation of crystal polymorphs, is manifested during the supramolecular assembly of synthetic polymers (79). Ostwald's rule affirms that the least stable polymorph would form first, followed by a subsequent rollover to the more stable form. We predict that for A$\beta$42, transition from the disordered ground state to the U-bend topology is likely to be faster, with the S-bend topology appearing only on longer observation time scales. This prediction, which follows from the inequality $\Delta G_{N_U^*} > \Delta G_{N_S^*}$ ($\Delta G_{N_U^*}$ and $\Delta G_{N_S^*}$ denoting the free energy difference of the U-bend and the S-bend motifs with respect to the disordered ground state), could be tested using kinetic simulations. It follows (Fig. 9) that formation of various polymorphic structures, which would not interconvert directly from one to another, could be under kinetic control (80).

## Materials and Methods

The A$\beta$40 and A$\beta$42 peptides are modeled using a modification of the recently introduced SOP-IDP model, which quantitatively and in unprecedented detail reproduces the scattering profiles of a diverse range of IDP sequences of varying lengths, sequence composition, and charge densities (52). We calculated the thermodynamic quantities using trajectories generated in low-friction Langevin dynamics simulations, which enhances conformational sampling (81). The simulations were used to calculate a number of observables for the A$\beta$40 and A$\beta$42 peptides, which were directly compared to experiments to validate the SOP-IDP model. Details of the simulations and analyses of the simulations may be found in *SI Appendix*.

**Data Availability.** The dataset and the analysis scripts are hosted by Zenodo (https://zenodo.org/record/3891924#.XwxyqJNKgY0). The parameters of the SOP-IDP force field used in this study are included in *SI Appendix*.

**ACKNOWLEDGMENTS.** We are grateful to Upayan Baul for assistance in the initial stages of this work. The comments of Robert Tycko and the encouragement of Tuomas Knowles are greatly appreciated. We also acknowledge Mauro Mugnai and Abhinav Kumar for several fruitful discussions. We acknowledge the Texas Advanced Computing Center for providing the necessary computing resources. This work was supported by a grant from the National Institutes of Health (GM-107703) and the Welch Foundation Grant F-0019 through the Collie-Welch chair.

1. R. Tycko, Amyloid polymorphism: Structural basis and neurobiological relevance. *Neuron* **86**, 632–645 (2015).
2. A. Aguzzi, C. Haass, Games played by rogue proteins in prion disorders and Alzheimer's disease. *Science* **302**, 814–818 (2003).
3. W. Qiang, W. Yau, J. X. Lu, J. Collinge, R. Tycko, Structural variation in amyloid-$\beta$ fibrils from Alzheimer's disease clinical subtypes. *Nature* **541**, 217–221 (2017).
4. V. H. Finder, R. Glockshuber, Amyloid-beta aggregation. *Neurodegener. Dis.* **4**, 13–27 (2007).
5. A. V. Maltsev, S. Bystryak, O. V. Galzitskaya, The role of $\beta$-amyloid peptide in neurodegenerative diseases. *Ageing Res. Rev.* **10**, 440–452 (2011).
6. E. D. Eanes, G. G. Glenner, Physical and chemical properties of amyloid fibers I. X-ray diffraction studies on amyloid filaments. *J. Histochem. Cytochem.* **16**, 673–677 (1968).
7. R. Nelson *et al.*, Structure of the cross-$\beta$ spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).
8. D. A. Kirschner, C. Abraham, D. J. Selkoe, X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-beta conformation. *Proc. Natl. Acad. Sci. U.S.A.* **1**, 503–507 (1986).
9. G. G. Glenner, C. W. Wong, Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* **120**, 885–890 (1984).
10. C. L. Masters *et al.*, Amyloid plaque core protein in Alzheimer disease and down syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4245–4249 (1984).
11. W. Qiang, W. M. Yau, J. X. Lu, J. Collinge, R. Tycko, Structural variation in amyloid-$\beta$ fibrils from Alzheimer's disease clinical subtypes. *Nature* **541**, 217–221 (2017).
12. S. B. Prusiner, Prions. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13363–13383 (1998).
13. W. P. Flavin *et al.*, Endocytic vesicle rupture is a conserved mechanism of cellular invasion by amyloid proteins. *Acta Neuropathol.* **134**, 629–653 (2017).
14. 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).
15. J. Nasica-Labouze *et al.*, Amyloid $\beta$ protein and Alzheimer's disease: When computer simulations complement experimental studies. *Chem. Rev.* **115**, 3518–3563 (2015).
16. I. M. Ilie, A. Caflisch, Simulation studies of amyloidogenic polypeptides and their aggregates. *Chem. Rev.* **119**, 6956–6993 (2019).
17. J. Roche, Y. Shen, J. H. Lee, J. Ying, A. Bax, Monomeric A$\beta$1–40 and A$\beta$1–42 peptides in solution adopt very similar Ramachandran map distributions that closely resemble random coil. *Biochemistry* **55**, 762–775 (2016).
18. L. Hou *et al.*, Solution NMR studies of the A$\beta$40 and A$\beta$42 peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *J. Am. Chem. Soc.* **126**, 1992–2005 (2004).
19. Y. Yan, C. Wang, A$\beta$42 is more rigid than A$\beta$40 at the C terminus: Implications for Abeta aggregation and toxicity. *J. Mol. Biol.* **364**, 853–862 (2006).
20. B. Tarus, J. E. Straub, D. Thirumalai, Structures and free-energy landscapes of the wild type and mutants of the A$\beta$21-30 peptide are determined by an interplay between intrapeptide electrostatic and hydrophobic interactions. *J. Mol. Biol.* **379**, 815–829 (2008).
21. G. Reddy, J. E. Straub, D. Thirumalai, Influence of preformed Asp23-Lys28 salt bridge on the conformational fluctuations of monomers and dimers of A$\beta$ peptides with implications for rates of fibril formation. *J. Phys. Chem. B* **113**, 1162–1172 (2009).
22. F. Massi, D. K. Klimov, D. Thirumalai, J. E. Straub, Charge states rather than propensity for $\beta$-structure determine enhanced fibrillogenesis in wild-type Alzheimer's $\beta$-amyloid peptide compared to E22Q Dutch mutant. *Prot. Sci.* **11**, 1639–1647 (2002).
23. D. K. Klimov, D. Thirumalai, Dissecting the assembly of $A\beta_{16-22}$ amyloid peptides into antiparallel $\beta$ sheets. *Structure* **11**, 295–307 (2003).
24. S. L. Bernstein *et al.*, Amyloid $\beta$-protein: Monomer structure and early aggregation states of A$\beta$42 and its Pro19 alloform. *J. Am. Chem. Soc.* **127**, 2075–2084 (2005).
25. A. Baumketner *et al.*, Structure of the 21-30 fragment of amyloid $\beta$-protein. *Protein Sci.* **15**, 1239–1247 (2006).
26. K. A. Ball *et al.*, Homogeneous and heterogeneous tertiary structure ensembles of amyloid-$\beta$ peptides. *Biochemistry* **50**, 7612–7628 (2011).
27. D. J. Rosenman, C. Wang, A. E. García, Characterization of A$\beta$ monomers through the convergence of ensemble properties among simulations with multiple force fields. *J. Phys. Chem. B* **120**, 259–277 (2016).
28. J. Lincoff, S. Sasmal, T. Head-Gordon, The combined force field-sampling problem in simulations of disordered amyloid-$\beta$ peptides. *J. Chem. Phys.* **150**, 104108 (2019).
29. Y. S. Lin, G. R. Bowman, K. A. Beauchamp, V. S. Pande, Investigating how peptide length and a pathogenic mutation modify the structural ensemble of amyloid beta monomer. *Biophys. J.* **102**, 315–324 (2012).

30. F. Meng *et al.*, Highly disordered amyloid-$\beta$ monomer probed by single-molecule FRET and MD simulation. *Biophys. J.* **114**, 870–884 (2018).
31. B. Tarus, J. E. Straub, D. Thirumalai, Dynamics of Asp23-Lys28 salt-bridge formation in A$\beta$10-35 monomers. *J. Am. Chem. Soc.* **128**, 16159–16168 (2006).
32. M. S. Li *et al.*, Factors governing fibrillogenesis of polypeptide chains revealed by lattice models. *Phys. Rev. Lett.* **105**, 218101 (2010).
33. P. Zhurvalev, G. Reddy, J. E. Straub, D. Thirumalai, Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. *J. Mol. Biol.* **426**, 2653–2666 (2014).
34. D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, A. E. García, A$\beta$ monomers transiently sample oligomer and fibril-like configurations: Ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* **425**, 3338–3359 (2013).
35. N. A. Eschmann *et al.*, Tau aggregation propensity engrained in its solution state. *J. Phys. Chem. B* **119**, 14421–14432 (2012).
36. H. B. Nam, M. Kouza, H. Zung, M. S. Li, Relationship between population of fibril-prone conformation in the monomeric state and oligomer formation times of peptides: Insights from all-atom simulations. *J. Chem. Phys.* **132**, 165104 (2010).
37. I. Kuperstein *et al.*, Neurotoxicity of Alzheimer's disease A$\beta$ peptides is induced by small changes in the A$\beta$42 to A$\beta$40 ratio. *EMBO J.* **29**, 3408–3420 (2010).
38. V. H. Finder, I. Vodopivec, R. M. Nitsch, R. Glockshuber, The recombinant amyloid-beta peptide A$\beta$1-42 aggregates faster and is more neurotoxic than synthetic A$\beta$1-42. *J. Mol. Biol.* **396**, 9–18 (2010).
39. G. Meisi *et al.*, Differences in nucleation behavior underlie the contrasting aggregation kinetics of the A$\beta$40 and A$\beta$42 peptides. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9384–9389 (2014).
40. J. D. Harper, P. T. Lansbury, Models of amyloid seeding in Alzheimer's disease and scrapie: Mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu. Rev. Biochem.* **66**, 385–407 (1997).
41. N. D. Lazo, M. A. Grant, M. C. Condron, A. C. Rigby, D. B. Teplow, On the nucleation of amyloid $\beta$-protein monomer folding. *Protein Sci.* **14**, 1581–1596 (2005).
42. M. Yang, D. B. Teplow, Amyloid $\beta$-protein monomer folding: Free-energy surfaces reveal alloform-specific differences. *J. Mol. Biol.* **384**, 450–464 (2008).
43. R. Roychaudhuri *et al.*, C-terminal turn stability determines assembly differences between A$\beta$40 and A$\beta$42. *J. Mol. Biol.* **425**, 292–308 (2013).
44. W. A. Walti *et al.*, Atomic-resolution structure of a disease-relevant A$\beta(1-42)$ amyloid fibril. *Proc. Natl. Acad. Sci. U.S.A.* **1**, E4976–E4984 (2016).
45. M. T. Colvin *et al.*, Atomic resolution structure of monomorphic A$\beta$42 amyloid fibrils. *J. Am. Chem. Soc.* **138**, 9663–9674 (2016).
46. Y. Xiao *et al.*, A$\beta$ (1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* **22**, 499–505 (2015).
47. N. Sgourakis, Y. Yan, S. McCallum, C. Wang, A. E. Garcia, The Alzheimer's peptides A$\beta$40 and 42 adopt distinct conformations in water: A combined MD/NMR study. *J. Mol. Biol.* **368**, 1448–1457 (2007).
48. K. A. Ball, A. H. Phillips, D. E. Wemmer, T. Head-Gordon, Differences in $\beta$-strand populations of monomeric A$\beta$40 and A$\beta$42. *Biophys. J.* **104**, 2714–2724 (2013).
49. S. Côté, P. Derreumaux, N. Mousseau, Distinct morphologies for amyloid beta protein monomer: A$\beta$1-40, A$\beta$1-42, and A$\beta$1-40(D23N). *J. Chem. Theory Comput.* **7**, 2584–2592 (2011).
50. B. Urbanc *et al.*, Molecular dynamics simulation of amyloid $\beta$ dimer formation. *Biophys. J.* **87**, 2310–2321 (2004).
51. W. Zheng, M. Y. Tsai, P. G. Wolynes, Comparing the aggregation free energy landscapes of amyloid beta(1-42) and amyloid beta(1-40). *J. Am. Chem. Soc.* **139**, 16666–16676 (2017).
52. U. Baul, D. Chakraborty, M. L. Mugnai, J. E. Straub, D. Thirumalai, Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. *J. Phys. Chem. B* **123**, 3462–3474 (2019).
53. D. Granata *et al.*, The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci. Rep.* **5**, 15449 (2015).
54. S. Wennmalm, V. Chmyrov, J. Widengren, L. Tjernberg, Highly sensitive FRET-FCS detects amyloid $\beta$-peptide oligomers in solution at physiological concentrations. *Anal. Chem.* **87**, 11700–11705 (2015).
55. M. Doi, S. F. Edwards, *The Theory of Polymer Dynamic* (Oxford University Press, 1988), vol. 73.
56. N. G. Sgourakis *et al.*, Atomic-level characterization of the ensemble of the A$\beta$ (1-42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *J. Mol. Biol.* **405**, 570–583 (2011).
57. A. R. Lam, D. B. Teplow, H. E. Stanley, B. Urbanc, Effects of the arctic (E 22$\rightarrow$G) mutation on amyloid $\beta$-protein folding: Discrete molecular dynamics study. *J. Am. Chem. Soc.* **130**, 17413–17422 (2008).
58. S. Vivekanandan, J. R. Brender, S. Y. Lee, A. Ramamoorthy, A partially folded structure of amyloid-beta(1-40) in an aqueous environment. *Biochem. Biophys. Res. Commun.* **411**, 312–316 (2011).
59. Y. Fezoui, D. B. Teplow, Kinetic studies of amyloid-protein fibril assembly. *J. Biol. Chem.* **277**, 36948–36954 (2002).
60. M. D. Kirkitadze, M. M. Condron, D. B. Teplow, Identification and characterization of key kinetic intermediates in amyloid $\beta$-protein fibrillogenesis. *J. Mol. Biol.* **312**, 1103–1119 (2001).
61. A. T. Frank, S. M. Law, L. S. Ahlstrom, C. L. Brooks, Predicting protein backbone chemical shifts from $C_\alpha$ coordinates: Extracting high resolution experimental observables from low resolution models. *J. Chem. Theory Comput.* **11**, 325–331 (2015).
62. O. O. Olubiyi, B. Strodel, Structures of amyloid $\beta$-peptides A$\beta$ (1-40) and A$\beta$ (1-42) as influenced by pH and a D-peptide. *J. Phys. Chem. B* **116**, 3280–3291 (2012).
63. L. Gu, S. Ngo, Z. Guo, Solid-support Electron Paramagnetic Resonance (EPR) studies of A$\beta$40 monomers reveal a structured state with three ordered segments. *J. Biol. Chem.* **287**, 9081–9089 (2012).
64. K. Brannstrom, T. Islam, L. Sandblad, A. Olofsson, The role of histidines in amyloid $\beta$ fibril assembly. *FEBS Lett.* **591**, 1167–1175 (2017).
65. J. McLaurin, P. E. Fraser, Effect of amino-acid substitutions on Alzheimer's amyloid-beta peptide-glycosaminoglycan interactions. *Eur. J. Biochem.* **267**, 6353–6361 (2000).
66. A. Baumketner, J. E. Shea, The structure of the Alzheimer amyloid $\beta$ 10-35 peptide probed through replica-exchange molecular dynamics simulations in explicit solvent. *J. Mol. Biol.* **366**, 275–285 (2007).
67. R. Cukalevski *et al.*, Role of aromatic side chains in amyloid $\beta$-protein aggregation. *ACS Chem. Neurosci.* **3**, 1008–1016 (2012).
68. W. P. Esler *et al.*, Point substitution in the central hydrophobic cluster of a human $\beta$-amyloid congener disrupts peptide folding and abolishes plaque competence. *Biochemistry* **35**, 13914–13921 (1996).
69. W. Han, Y. D. Wu, A strand-loop-strand structure is a possible intermediate in fibril elongation: Long time simulations of amyloid-beta peptide (10-35). *J. Am. Chem. Soc.* **127**, 15408–15416 (2005).
70. L. Gremer *et al.*, Fibril structure of amyloid-$\beta(1-42)$ by cryo–electron microscopy. *Science* **358**, 116–119 (2017).
71. T. Luhrs *et al.*, 3D structure of Alzheimer's amyloid-$\beta(1-42)$ fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17342–17347 (2005).
72. S. K. Maji, J. J. Amsden, K. J. Rothschild, M. M. Condron, D. B. Teplow, Conformational dynamics of amyloid $\beta$-protein assembly probed using intrinsic fluorescence. *Biochemistry* **44**, 13365–13376 (2005).
73. N. L. Fawzi *et al.*, Structure and dynamics of the Abeta(21-30) peptide from the interplay of NMR experiments and molecular simulations. *J. Am. Chem. Soc.* **130**, 6145–6158 (2008).
74. M. Kollmer *et al.*, Cryo-EM structure and polymorphism of A$\beta$ amyloid fibrils purified from Alzheimer's brain tissue. *Nat. Commun.* **10**, 4760 (2019).
75. J. X. Lu *et al.*, Molecular structure of $\beta$-amyloid fibrils in Alzheimer's disease brain tissue. *Cell* **154**, 1257–1268 (2013).
76. M. Fandrich, M. A. Fletcher, C. M. Dobson, Amyloid fibrils from muscle myoglobin. *Nature* **410**, 165–166 (2001).
77. K. L. Sciaretta, D. J. Gordon, A. T. Petkova, R. Tycko, S. C. Meredith, A$\beta$40-Lactam(D23/K28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry* **44**, 6003–6014 (2005).
78. P. Neudecker *et al.*, Structure of an intermediate state in protein folding and aggregation. *Science* **20**, 362–366 (2012).
79. A. Levin *et al.*, Ostwald's rule of stages governs structural transitions and morphology of dipeptide supramolecular polymers. *Nat. Commun.* **5**, 5219 (2014).
80. R. Pellarin, P. Schuetz, E. Guarnera, A. Calfisch, Amyloid fibril polymorphism is under kinetic control. *J. Am. Chem. Soc.* **132**, 14960–14970 (2010).
81. J. D. Honeycutt, D. Thirumalai, The nature of folded states of globular proteins. *Biopolymers* **32**, 695–709 (1992).