doi:10.1016/j.jmb.2010.08.003 J. Mol. Biol. (2010) 403, 148–165

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Micelle-Like Architecture of the Monomer Ensemble of Alzheimer's Amyloid-β Peptide in Aqueous Solution and Its Implications for Aβ Aggregation

Andreas Vitalis* and Amedeo Caflisch*

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Received 22 June 2010;
received in revised form
27 July 2010;
accepted 2 August 2010
Available online
13 August 2010

Edited by D. Case

Keywords:
Alzheimer's disease;
Monte Carlo simulations;
conformational equilibria;
disorder;
hydrophobic interactions

Aggregation of amyloid-β (Aβ) peptide, a 39- to 43-residue fragment of the amyloid precursor protein, is associated with Alzheimer's disease, the most common form of dementia in the elderly population. Several experimental studies have tried to characterize the atomic details of amyloid fibrils, which are the final product of Aβ aggregation. Much less is known about species forming during the early stages of aggregation, in particular about the monomeric state of the Aβ peptide that may be viewed as the product of the very first step in the hypothesized amyloid cascade. Here, the equilibrium ensembles of monomeric Aβ alloforms $A\beta_{1-40}$ and $A\beta_{1-42}$ are investigated by Monte Carlo simulations using an atomistic force field and implicit solvent model that have been shown previously to correctly reproduce the ensemble properties of other intrinsically disordered polypeptides.

Our simulation results indicate that at physiological temperatures, both alloforms of Aβ assume a largely collapsed globular structure. Conformations feature a fluid hydrophobic core formed, on average, by contacts both within and between the two segments comprising residues 12–21 and 24–40/42, respectively. Furthermore, the 11 N-terminal residues are completely unstructured, and all charged side chains, in particular those of Glu22 and Asp23, remain exposed to solvent. Taken together, these observations indicate a micelle-like† architecture at the monomer level whose implications for oligomerization, as well as fibril formation and elongation, are discussed. We establish quantitatively the intrinsic disorder of Aβ and find the propensity to form regular secondary structure to be low but sequence specific. In the presence of a global and unspecific bias for backbone conformations to populate the β-basin, the β-sheet propensity along the sequence is consistent with the arrangement of the monomer within the fibril, as derived from solid-state NMR data. These observations indicate that the primary sequence partially encodes fibril structure, but that fibril elongation must be thought of as a templated assembly step.

**Corresponding authors.* E-mail addresses: a.vitalis@bioc.uzh.ch; caflisch@bioc.uzh.ch.

Abbreviations used: Aβ, amyloid β; IDP, intrinsically disordered protein; CHC, central hydrophobic core; REMC, replica-exchange Monte Carlo; PMF, potential of mean force; SAV, solvent-accessible volume; PDB, Protein Data Bank.

† The terms "micelle-like architecture" and "micelle-like" used throughout this article refer to the average conformation of an individual polypeptide chain and do not imply the assembly of several molecules, unless otherwise noted.

Introduction

The famous amyloid cascade hypothesis[1] states that the pathogenesis of Alzheimer's disease is triggered by the aggregation of a fragment of the human amyloid precursor protein, termed amyloid-β (Aβ). Its two most common alloforms, $A\beta_{1-40}$ and $A\beta_{1-42}$, show significantly different pathogenicities *in vivo*[2] and aggregation propensities *in vitro*,[3] despite the addition of only two residues (IA) at the C-terminus. As the "educt" of a pathogenic polymerization reaction, the monomeric form of both peptides continues to demand much attention despite the difficulties associated with obtaining interpretable results experimentally.[4] The reason for the latter lies in the peptides' intrinsic disorder and their strong aggregation propensity. Disorder will often produce a lack of signal in ensemble measurements,[5] and the tendency to aggregate[6] requires specialized assays and controls to ensure that one is in fact studying the unperturbed monomer ensemble.[7]

NMR data suggest that monomeric Aβ exists as disordered coils in solution, although it appears possible to bias the ensemble toward certain secondary structure elements by changing solution conditions and/or the oxidation state of Met35.[8–12] There seems to be no consensus on assigning residual secondary structure propensities, although a careful synthesis of all experimental data may achieve that. At high enough concentration, Aβ forms distinct oligomers with alloform-specific distributions.[13–15] Similar to the debate on the monomer ensemble, the interpretation of *in vitro* aggregation data and conclusions about mechanisms and pathways remain partially unanswered. Under most conditions, Aβ aggregation is described as a nucleation-dependent process,[16–18] the underlying complexity of which may easily be masked by the specific technique employed to monitor aggregation or special solution conditions.[19,20] A prominent end product of the aggregation process *in vitro* are amyloid fibrils similar to those found in extracellular plaques extracted from the brains of Alzheimer's disease patients. Solid-state NMR measurements, combined with many other techniques, have been used extensively to obtain structural models for mature Aβ fibrils.[21,22] They appear to be characterized by a hairpin-like architecture with intermolecular β-hydrogen bonds, excluding the N-terminal residues. The fiber axis runs perpendicular to the hairpin axis, and the turn is formed around residues 25–30. The N-terminal portion of the peptides is usually not resolved. As polymorphism is prominent[23–25] and as aggregates at the oligomer level appear to be much more neurotoxic,[26,27] atomic-resolution structures of amyloid fibrils may not prove to be as useful as once hoped.

Admittedly, it seems difficult at first sight to justify a further *in silico* study of the properties of monomeric Aβ.[28] Moore's law has enabled simulations to now attempt to describe oligomerization events[29] or to assess the stabilities of various supramolecular assemblies.[30,31] Nonetheless, computer simulation may hold particular promise at the monomer level by providing atomic-resolution access to conformational ensembles consistent with, ideally, all of the ensemble data measured experimentally under similar conditions. In fact, molecular simulations at an appropriate resolution may be the only route to obtain such information within the next decade. Throughout the last 5 years, several groups have carefully studied the full-length alloforms of Aβ *in silico*.[32–35] However, the ensembles were typically analyzed with a focus on structural details that appear to differ even though an approximate agreement with at least one set of experimental data was generally shown. This points toward two major problems with simulation work on such a complex problem: (i) statistical precision and (ii) model bias. While model bias may account for some of the discrepancies, it need not even be considered in the absence of data obtained with the necessary precision. As a specific example, we may consider the intramolecular contact analyses performed by Yang and Teplow[32] and Sgourakis *et al.*[33] In either case, only a single replica-exchange run was performed, and the only error estimates presented are those on predicted NMR data. It is therefore impossible to extract quantitative information: In essence, we do not know how to interpret the differences in those contact maps, and we also do not know which data set to trust more. These issues are confounded by the fact that data such as contact maps may be presented only over subensembles,[33] or that inherently non-transferable metrics of analysis such as principal components are used.[32]

In this study, we choose a simple enough model and an appropriate methodology to eliminate concerns about the precision of our results. Specifically, we perform simulations many times in parallel with completely randomized and independent starting conditions, enabling us to distinguish robust trends from noise. Only error estimates obtained by comparing completely independent simulations to each other guarantee a complete lack of correlation between the data, and we indicate such errors in nearly all plots. Completely uncorrelated data are typically not considered when, for example, obtaining errors from block averages even when using replica-exchange molecular dynamics. Our model of choice to describe effective interactions in an aqueous environment is the ABSINTH implicit solvation model.[36] This model has been used successfully in the past to study another aggregation-prone intrinsically disordered protein (IDP), viz. polyglutamine, both as a homopolymer[37,38]

and attached to a heteropolymeric N-terminal domain derived from the protein Huntingtin.[39] Mao *et al.* showed that measured translational diffusion coefficients of arginine-rich peptides are consistent with simulation data employing the ABSINTH model and indicate a net-charge-induced globule-to-coil transition.[40] Lastly, the original publication presented data showing that the temperature-induced order-to-disorder transition of several polypeptides capable of folding is well described by the ABSINTH model.[36] Taken together, these results indicate that the model is well suited to describe an intrinsically disordered system such as Aβ, and that it is unlikely to overemphasize specific structural biases or to produce artifactual results due to an imbalanced description of the relative strengths of different types of interactions.

We investigate the conformational distributions of monomeric Alzheimer's Aβ peptides in both common alloforms, $A\beta_{1-40}$ and $A\beta_{1-42}$, using geometrical descriptors and cluster analysis. We show throughout that the sampled ensembles are consistent with important pieces of experimental data on the full-length peptides. Using polymeric metrics, we show Aβ to be a heterogeneous ensemble of collapsed states in water. As a consensus "mean" topology, we establish that Aβ preferentially adopts a micelle-like architecture driven by dominant hydrophobic stretches [the central hydrophobic core (CHC),[41] which encompasses residues 17–21 (i.e., LVFFA), and C-terminal residues 31–36 (i.e., IIGLMV)]. Monomeric Aβ resembles a canonical lipid micelle by the sequestration of hydrophobic side chains (hydrophobic "tails") away from solvent that is accompanied by full solvent exposure of charged side chains (polar "head groups"). The former constitute the hydrophobic core whose fluidity we establish by quantifying Aβ's intrinsic disorder. Highly heterogeneous tertiary contacts are observed, and the ensemble topology is more akin to a fluid micelle rather than to a folded protein. Next, we analyze secondary structure propensities at thermal equilibrium and in the presence of an ordering bias. The bias serves to amplify sequence-specific propensities for the formation of β-secondary structure, which we illustrate by comparing our data to experimentally determined structures of Aβ. The data reveal that the generally low β-propensity at equilibrium nonetheless shadows experimental observations in the presence of a suitable template in sequence-specific fashion.

The rest of the article is structured as follows. In Results, we describe the results of all of our simulations. We discuss the relevance of our findings in the context of the experimental literature and attempt to project their meaning for the aggregation process of Aβ. Future directions emerging immediately from this contribution are touched upon. A brief description of the methodology that is largely consistent with prior work concludes the article.

Results

It is of paramount importance to continuously assess the congruence of our simulation results to experimental data that are probing different length scales. As outlined in Materials and Methods, we employ the ABSINTH implicit solvation model[36] to approximate the effects of water while representing the inorganic ions of a low-ionic-strength electrolyte solution explicitly. As outlined in the Introduction, recent work employing the ABSINTH model[36–40] suggests that its application to an aggregation-prone IDP with a low net charge, such as Aβ, will be adequate.

Both alloforms of Aβ prefer collapsed conformations at physiological temperatures

Conclusive experimental data on monomeric Aβ are difficult to obtain due to the peptides' intrinsic disorder and their inherent propensity to form oligomers and aggregates, which may contaminate results. Zhang *et al.* managed to obtain an NMR ensemble for $A\beta_{10-35}$-NH_2 at 283 K showing that this peptide adopts a collapsed structure in water devoid of canonical secondary structure.[42] $A\beta_{10-35}$-NH_2 excludes Aβ's N-terminus, which is known to be disordered and solvent accessible at least in the fibril,[43,44] as well as predominantly hydrophobic residues at the C-terminus including two consecutive glycine residues. The implied global result of collapse, rather than the specific structure, is corroborated by most *in silico* studies on long enough fragments of Aβ.[45,46] Figure 1a shows that at temperatures and at an ionic strength (130 mM) comparable to *in vitro* experiments our simulations predict collapsed ensembles, as evidenced by the plateauing observed in a plot of the mean residue–residue distance as a function of sequence separation. The scaling of internal distances is tied to a rigorous polymer theoretical footing and allows us to conclude that the chains do not adopt extended conformations under these conditions, since such fractal objects would result in polynomial scaling behavior ($\langle R_{ij} \rangle \sim |i-j|^{\nu}$, where ν is the scaling exponent) rather than a plateau in this analysis.[48] As expected, swelling is observed with increasing temperature, but the molecules predominantly reside in the globule phase for all of the conditions studied here. This is corroborated by the data in (c), which show identical plots for the crystal structures of three different proteins. There are some subtle differences reflecting the unique topology in each case, such as the undulations for the all-β-sheet protein tendamistat that are caused by repeating

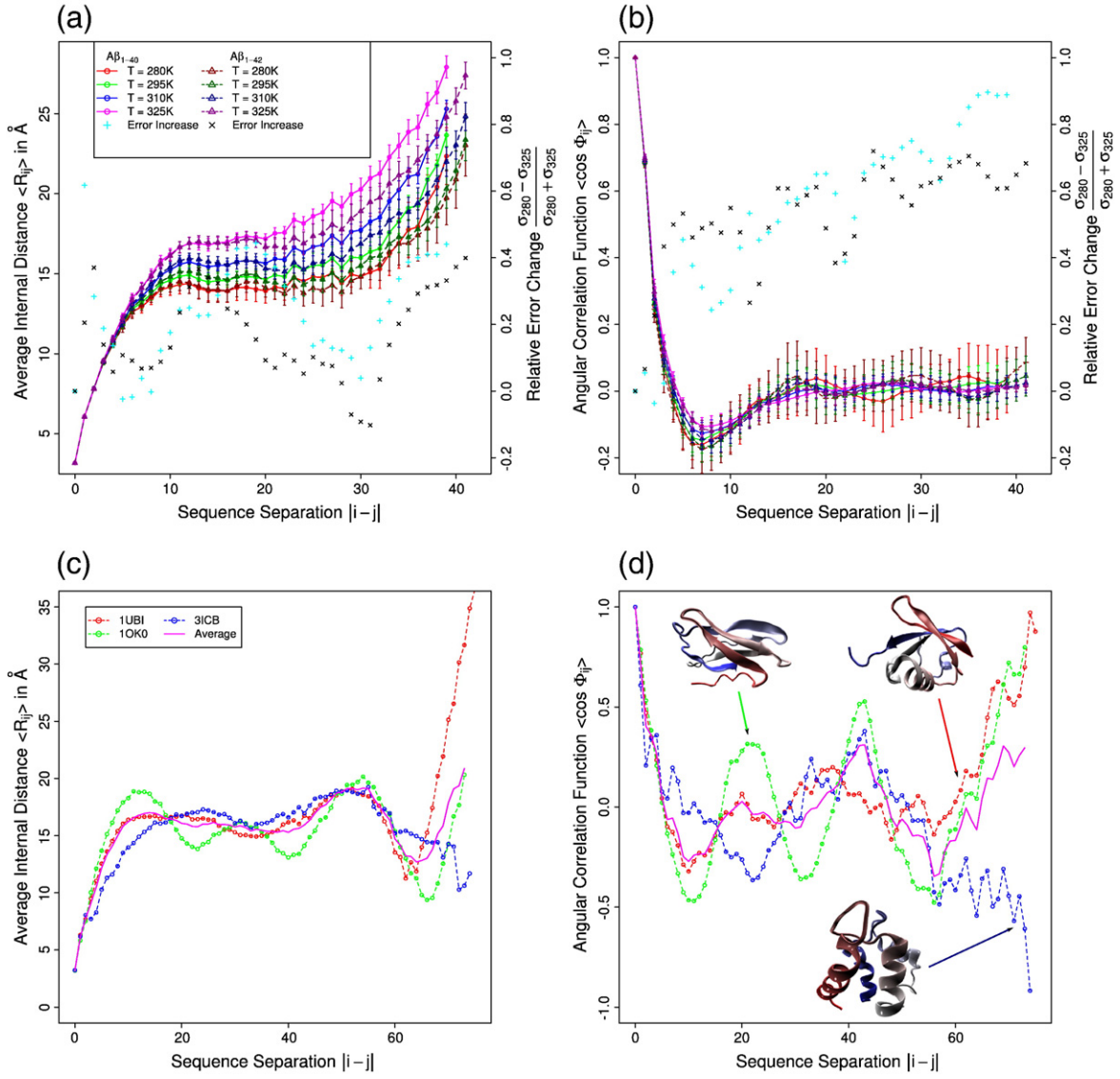


Fig. 1. Average polymeric properties of Aβ. Equilibrium data for four different temperatures and both alloforms are shown. (a) The mean interatomic distance between residues plotted as a function of sequence separation (left *y*-axis). For collapsed globules, this plot exhibits a prominent plateau region for all, but very small and maximal sequence separations. (b) The average angular correlation function between the N→C vectors of pairs of residues plotted as a function of sequence separation. Negative values indicate a requirement for the chain to turn on itself (i.e., usually report on collapsed globules). As in all other plots, lines are only a guide to the eye, and standard errors were obtained as described in the text. In both panels, the normalized relative changes in standard errors from 280 K to 325 K are shown with symbols only. (c and d) Data analogous to (a) and (b) for three different folded proteins with lengths from 74 to 76 residues and three different topologies (ubiquitin, PDB ID 1UBI; tendamistat, PDB ID 1OK0; calbindin, PDB ID 3ICB), and their average. Note the different *y*-axis scaling compared to the top row. We wish to emphasize the similarity between the data in (a) and the data in (c), and, conversely, the much more pronounced differences observed in (d) between the different protein topologies and compared to the ensemble-averaged topology of a disordered system seen in (b). To illustrate the sources of the signals in (c) and (d), we added cartoon representations[47] of the three proteins to (d). The coloring scheme indicates the N-terminus (red) and the C-terminus (blue).

antiparallel β-strand arrangements. Nonetheless, the overall plateau between sequence separations of 10–60 residues is clearly discernible and reports on the globular shape and compactness common to all three proteins.

Danielsson *et al.* noted that the estimated persistence length for $A\beta_{1–40}$ drops from eight to less than three residues if the temperature is raised from 3 °C to 18 °C.[49] This estimate is based on NMR relaxation data and was used to conclude that at

physiologically relevant temperatures the peptide forms a "random coil." Figure 1b shows the ensemble-averaged angular correlation function for both peptides observed in our simulations. When assuming negative values, this correlation function reports on the tendency of the backbone to turn on itself. Conversely, swollen flexible coils would be characterized by strictly positive values and exponential decay behavior. All curves indicate the presence of disordered and collapsed globules. This leads to the characteristic dip over separations of 5–10 residues and a cancellation of correlations beyond that. The disorder becomes apparent by comparison to Fig. 1d, which shows the angular correlation function for the three aforementioned proteins. Here, we observe three disparate curves indicative of the proteins' unique topologies. Long-distance correlations do not cancel as they do for disordered Aβ in Fig. 1b. Not surprisingly, the average taken over the three proteins (magenta line) resembles the Aβ data the most. Indirect evidence of the NMR relaxation data can be found in the error estimates for these data, which increase strongly from the highest temperature to the lowest temperature studied. For comparison, standard errors on the data in Fig. 1a increase with decreasing temperature as well, but not nearly to such an extent. We can conclude that the low-temperature relaxation behavior in NMR is mirrored in a difficulty to interconvert between different collapsed conformations. We also note that very short persistence length estimates are more consistent with collapsed states that rapidly interconvert, rather than with extended states.[50] Lastly, it should be pointed out that one could attempt to extract persistence lengths from Fig. 1b, but that such data would not be directly comparable to the NMR-based estimates because the former represent structural estimates and the latter represent kinetic estimates.

The ensemble exhibits a preferred topology with a hydrophobic core formed by distal contacts

The analysis thus far has suggested that the collapsed states adopted by $A\beta_{1-40}$ and $A\beta_{1-42}$ are random in nature. Figure 2 shows ensemble-averaged contact maps for both peptides revealing that this statement has to be refined. From the contact maps, we observe that local contacts along the chain are found throughout the entire sequence, but are most pronounced at the C-terminus, presumably due to the large number of glycine residues in that region, which facilitate tight turns. Extensive nonlocal contacts are observed between the C-terminal segment (residues 24–40/42) and residues 12–21. These dominant distal contacts indicate the presence of a relatively well-defined, albeit fluctuating, hydrophobic core. The observed "fluidity" implies that not all hydrophobic residues will form a connected network at all times, but may be broken up into separate clusters. However, the overall globular architecture constrains buried

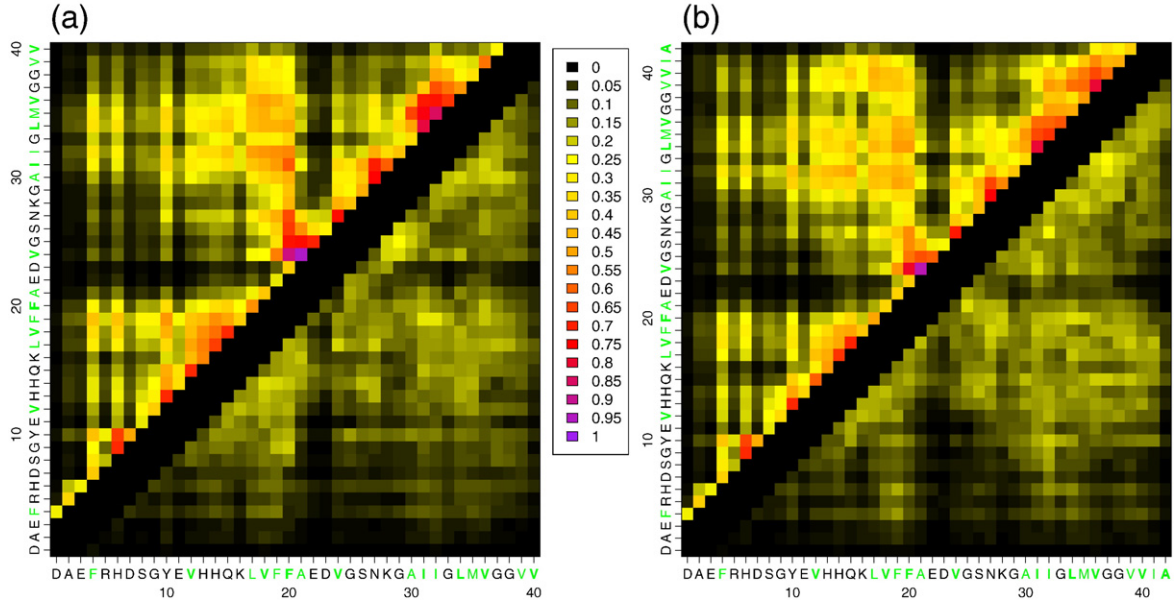


Fig. 2. Average residue–residue contact maps for Aβ. Equilibrium data at 310 K are shown in the upper-left half matrices for $A\beta_{1-40}$ (a) and $A\beta_{1-42}$ (b). A contact was deemed to be present if any atom–atom distance between a specific pair of residues dropped below 5.5 Å. Standard errors are shown in the lower-right half matrices. Almost all of the prominent long-range contacts with frequencies of occurrence exceeding 30% appear to be statistically significant. In the axis labels, hydrophobic residues are highlighted in green.

residues to reside in spatial proximity given the comparatively small size of the globules (see Fig. 3). The two aforementioned regions were shown to be the only ones exhibiting significant binding in an assay testing the interaction of 10-residue fragments derived from $A\beta_{1-40}$ with full length $A\beta_{1-40}$.[51] This is in direct support of the results presented here. Furthermore, residues 12–21 showed the highest tendency for ordered in-register aggregation in previous molecular dynamics simulations with a solvent-accessible surface area model of solvation.[52] The N-terminal residues participate in long-range contacts more sporadically, indicating that this part of the peptide will often be "dissociated" from the rest. Furthermore, residues 22 and 23 show negligible propensity to engage in intramolecular contacts for either peptide, implying that they remain solvent exposed throughout. This may pose a form of topological constraint to the ensemble due to the requirement of those central residues being exposed on the surface of the globule.

It would be misleading, however, to suggest that the observed ensemble-averaged topology is the result of each conformation sharing a set of identical features. Instead, disorder prevails, and the dissimilarity between conformations is generally large. Figure 4a shows the distribution of pairwise distances within a single replica-exchange Monte Carlo (REMC) run measured by two structures' root-mean-square deviation (RMSD). At 310 K, the number of self-similar (<5 Å) structures is exceptionally small, and the RMSD values appear to be distributed normally. This shows that many different conformations are visited during a single run (i.e., that the sampling quality is good at this temperature). Conversely, at 280 K, substantial density appears for pairs of structures with small RMSD values (<5 Å). While the ensemble must still be considered primarily disordered, it appears as if basin interconversion is dramatically slowed at this temperature. This is consistent with the prior analysis (see error analysis in Fig. 1) and implies a much increased difficulty in obtaining converged ensemble averages. Lastly, pairwise RMSD distributions are generally not significantly different between the two peptides.

Figure 4b–d attempt to corroborate the conclusions drawn based on the analysis of intramolecular contacts. Figure 4b shows a comparison of the radii of gyration of two subsets of peptide heavy atoms: all hydrophobic side chains from residue 12 to residue 40/42 *versus* all hydrophilic side chains from residue 1 to residue 28. Both subsets comprise a roughly equal number of residues and sequence length, and one could reasonably expect the distributions to be overlapping in the absence of a solvent environment. The data in Fig. 4 show that the solvent has a strong ordering effect. At both plotted temperatures and for both chains, the

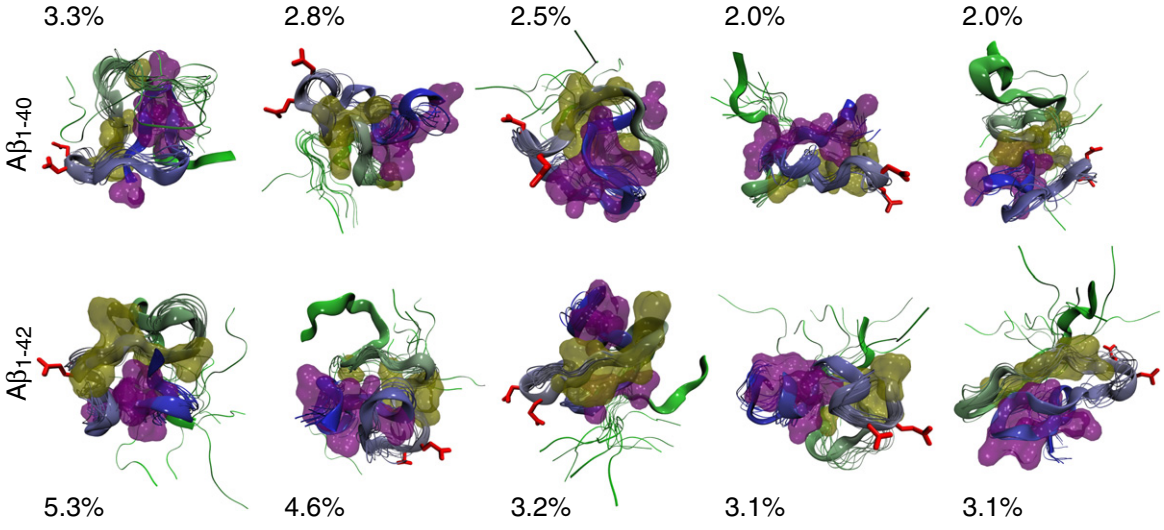


Fig. 3. Central structures of clusters obtained at 280 K (see Materials and Methods for details). The top row shows the five most populated clusters for $A\beta_{1-40}$, and the bottom row shows the same for $A\beta_{1-42}$. The fraction of the ensemble they encompass is indicated for each image. The backbone of the central structure is displayed as broad ribbons ranging in color from green (N-terminus), via gray, to blue (C-terminus). Transparent surface representations of hydrophobic side-chain heavy atoms are added for residues 12–21 (yellow) and residues 30–40/42 (magenta), respectively. The side-chain heavy atoms of residues 22 and 23 are shown in red licorice representation. Additionally, the backbone of nine other structures from the same cluster (chosen at uniform interval from the ordered list of snapshots for the respective cluster) is shown as narrow ribbons. This is meant to illustrate the span of the clusters. Note that the alignment ignores the 11 N-terminal residues, but that meaningful alignment and clustering were typically not possible when including them. Rendering was performed using VMD.[47]

hydrophobic side chains have a much smaller radius of gyration, indicating their preferential orientation toward the "inside" of the globules (i.e., their tendency to form a loose and dynamic core). Conversely, the nonhydrophobic side chains seem much more dispersed. Interestingly, this trend is not weakened with increasing temperature even though the peptides overall swell (see Fig. 1a). Taken together, these data indicate that in terms of side-chain orientation Aβ assumes micelle-like conformations in water, much similar, at first glance, to the partitioning of side chains in folded proteins. However, Fig. 4c and d show that there is an important difference.

Here we plot the average solvent accessibility per residue, as defined within the ABSINTH

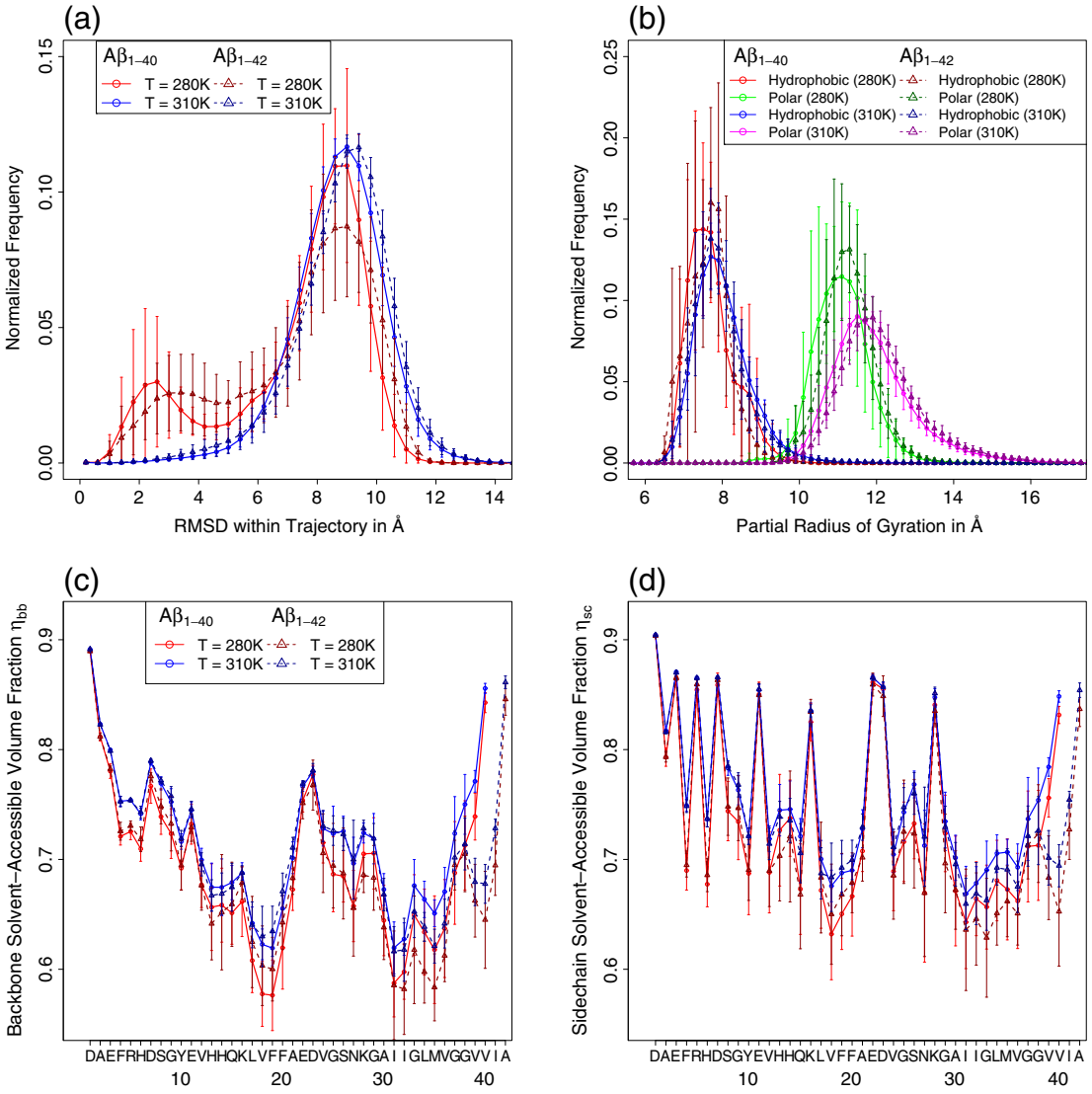


Fig. 4. Structural heterogeneity, micelle-like architecture, and solvent accessibility of Aβ. (a) Histograms of pairwise C^{α} RMSD values at two different temperatures for both alloforms. Given a specific temperature, alloform, and REMC run, 900 evenly spaced snapshots were compared to every second snapshot from the same trajectory yielding 4.05×10^5 pairs for comparison. The resultant histograms for all 12 independent REMC runs were combined and are shown in (a). This plot is useful in extracting a meaningful upper bound for the self-similarity of conformations encountered in the simulations. (b) Histograms of partial radii of gyration: (i) all heavy atoms from the side chains of hydrophobic residues in the range 12–40/42, and (ii) all heavy atoms from the side chains of nonhydrophobic residues in the range 1–28. The sets were chosen to be as large as possible while encompassing approximately the same sequence length and number of side chains. Note that standard errors on histogram data ignore the implicit normalization constraint. (c and d) The mean SAV fractions η[36] averaged over all backbone atoms (c) and side-chain atoms (d) of each residue, respectively. Data are presented for two temperatures and both chains. The legend in (c) applies to (d) as well.

implicit solvation model.[36] Irrespective of whether side-chain (Fig. 4d) or backbone atoms (Fig. 4c) are considered, the least solvent-accessible parts of Aβ are the CHC (L17-A21) and the C-terminal hydrophobic region (in particular, I31-V36). This corroborates the contact analyses (see Fig. 2), which suggested those two regions as participating in dominant long-range contacts. Unlike in a folded protein, however, the solvent accessibility remains significantly larger (>55%) than the smallest possible value of ~26% for close packing of spheres.[36] This is true for both peptides, and the values increase smoothly with temperature. This incomplete hydrophobic sequestration from solvent is attributable to the peptides' intrinsic disorder (Fig. 4a), which is illustrated qualitatively next.

Figure 3 shows the central structures of the five most populated clusters for either alloform from an analysis described in detail in Materials and Methods. The essence of the clustering was to find molecular topologies that are representative and highlight the features established quantitatively in Figs. 1, 2, and 4. We chose the lowest temperature (280 K) and restricted the alignment set primarily by excluding the highly flexible N-terminal region. Figure 3 illustrates several points made previously:

(1) There are extensive contacts within and between the two dominant hydrophobic segments in the chains. However, the details of these contacts are highly variable, and not all contacts observed in Fig. 2 are ever observed simultaneously (we wish to point out that the contact maps at 280 K are qualitatively indistinguishable from those at 300 K).
(2) This implies that occasionally few smaller clusters are formed instead of a single larger one. Figure 3 overemphasizes this trend by only highlighting the side chains of canonically hydrophobic residues while ignoring neutral, polar residues. The latter are commonly found not only on the surface but frequently also on the inside of proteins, and may constitute part of the core.
(3) Hydrophobic side chains often form the core of the globules and are sequestered from solvent, but individual residues or patches remain (partially) solvent exposed.
(4) Residues E22 and D23 are always solvent exposed, and their backbone often forms a tight turn.
(5) The N-terminus is frequently dissociated from the globular remainder of Aβ, potentially more so for $A\beta_{1-42}$ than for $A\beta_{1-40}$.
(6) Disorder is emphasized by the fact that even at 280 K self-similarity is so low that lenient coarse graining finds clusters that each represent only a few percentages of the data. A similar coarse-graining attempt was completely unsuccessful at 310 K, in agreement with Fig. 4a.

The overall secondary structure content is low, but propensities are sequence specific

Riek *et al.* showed that the two alloforms $A\beta_{1-40}$ and $A\beta_{1-42}$ are not significantly different from each other in solution NMR analyses, and that both are devoid of canonical secondary structure and may be classified as disordered.[8] While parts of these conclusions have been challenged,[49] at physiological temperatures, it seems more or less indisputable that highly ordered states are insignificantly populated as determined by CD[53] or NMR.[8,10]

Figure 5a shows β-sheet propensities as defined by the frequency of being assigned to state "E" in the DSSP assignments of Kabsch and Sander for $A\beta_{1-40}$.[54] These hydrogen-bond-based analyses reveal that several stretches of the peptide have negligible β-propensity. The two stretches exhibiting the highest β-contents are those from residues 12–21 and 30–38/40, which coincide very well with the hydrophobic stretches identified via the analysis of intramolecular contacts. β-Content generally increases with decreasing temperature, consistent with a higher degree of order at lower temperatures. Figure 5b shows the same data for $A\beta_{1-42}$, and it becomes apparent that—in congruence with Riek *et al.*'s NMR data—the differences between the two sets of data are subtle. The most prominent change is introduced, not surprisingly, by the two added residues at the C-terminus, which appear to enable an elongation of β-secondary structure elements beyond the two glycine residues in positions 37 and 38. Some additional conclusions may be drawn. First, results from the contact maps suggested that the N-terminus and residues E22 and D23 are not commonly part of long-range intramolecular contacts. Equivalent conclusions may be drawn from the DSSP data that predict a similar exclusion from participation in β-hydrogen bonds. The latter are usually established via long-range contacts. Second, glycine residues act as strong breakers of β-secondary structures, evidenced by dips in β-propensities around glycines located toward the C-terminus.

Figure 5c and d show analogous analyses for all right-handed helical DSSP assignments. Interestingly, for both peptides, the regions with α-propensity overlap mostly with those shown to exhibit the highest β-propensities. The α-content is overall fairly low, emphasizing again the disordered nature of the ensemble, and it generally decreases with increasing temperature. Unlike β-secondary structure, α-helices

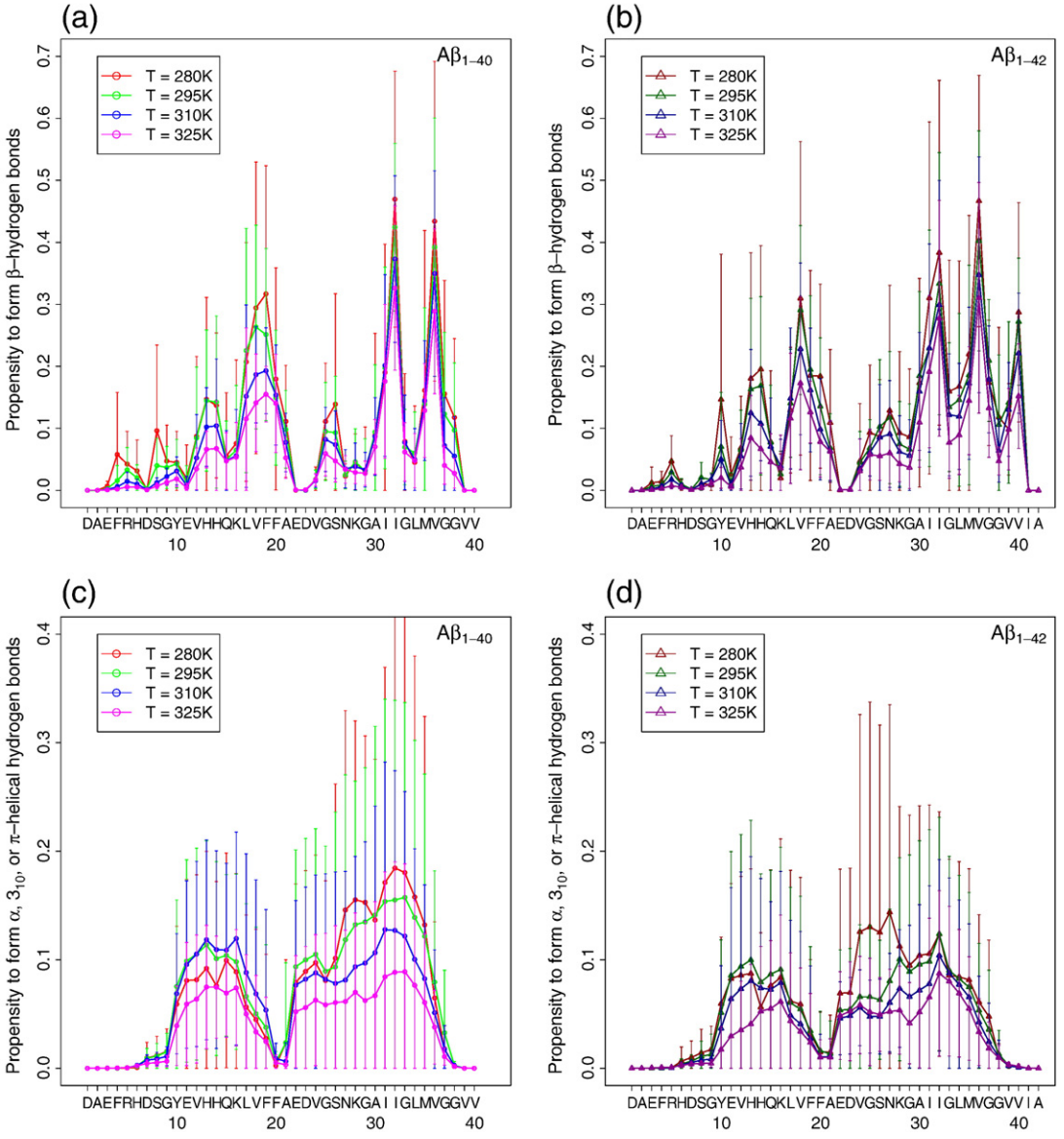


Fig. 5. Secondary structure propensities at thermal equilibrium. Data for all temperatures are shown for $A\beta_{1-40}$ (a and c) and $A\beta_{1-42}$ (b and d). (a and b) Frequencies of residues that are part of β-secondary structure elements, as defined by DSSP (state "E").[54] (c and d) Analogous data for all right-handed helices with $i \rightarrow i+3$, $i \rightarrow i+4$, or $i \rightarrow i+5$ hydrogen-bond patterns (state "G," "H," or "I," respectively). Right-handed α-helices (i.e., state "H") make, by far, the largest contribution to the signals observed in (c) and (d). Note that these data are exceptionally noisy, indicating substantial heterogeneity between individual runs.

seem to be able to incorporate glycine residues without too much trouble and to span residues 22–29 as well. Conversely, the N-terminus does not seem to be able to attain local order. It is worth pointing out that the results presented in Fig. 5c and d are consistent, in this regard, with NMR structures obtained for $A\beta_{1-28}$ peptide mutants in a helix-inducing environment,[55] as the latter show similarly disordered N-termini (residues 1–14), while the rest of the peptides are predominantly α-helical.

Propensities to form β-secondary structure along the sequence shadow the arrangement in fibrillar aggregates

Vitalis *et al.* previously quantified the global β-propensity of another aggregation-prone IDP, polyglutamine.[38] They employed an uninformed reaction coordinate, f_β, measuring the total β-content of the chain and performed umbrella sampling simulations to obtain the potential of mean force (PMF) along f_β, which is defined purely

in terms of dihedral angles. Here, we repeat this approach for Aβ to ascertain that the low β-content observed in equilibrium simulations is not the result of sampling difficulties.

Figure 6a shows the resultant PMF for both alloforms, which is mostly featureless and in global agreement for all runs performed. This implies statistical robustness and the insignificance of differences between the two peptides, in agreement with NMR studies.[8,33] Since the results suggest a largely harmonic basin in f_β space, we computed synthetic model PMFs based on entropy alone. The broken lines in Fig. 6a were obtained by considering $-k_bT\ln p_b$ as a suitable model. Here, p_b is the binomial distribution with a probability of 0.2 taken as the estimated mean of the actual PMFs and a size of 41 or 43 trials, respectively. This distribution reports simply on the entropic chance of finding anywhere from 0 to 40/42 residues in the β-basin. By normalization of the abscissa to describe β-content, the broken lines in Fig. 6 are made directly comparable to the actual PMFs. Comparison reveals that a large part of the data can in fact be described reasonably well by such a simple model. This indicates either the relative weakness of cooperative effects in the formation of β-rich structures for these peptides, or a compensatory behavior where the polymerically unfavorable swelling at high f_β values counteracts the favorable formation of β-hydrogen bonds.

In Fig. 6b, we plot the biased f_β histograms for both peptides and both runs at 310 K. This serves to illustrate that the umbrella sampling and the subsequent PMF determination via the weighted histogram analysis method are not marred by systematic errors. Overlap between neighboring windows is large throughout, and the histograms themselves are smooth. The statistical precision of the analysis is, of course, most strongly supported by the agreement between runs (see Fig. 6a). Figure 6 emphasizes that, under physiological conditions, the monomer faces a large free-energy barrier in converting into a structure that is high in β-content, such as those consistent with structural data on Aβ fibrils.[21,22] This, in itself, is of course not surprising: Aβ aggregation is well known to be a complex nucleation-dependent process with characterizable intermediates that are low in β-content.[17,27,56] We therefore have no reason to assume that structural transitions at the monomer level are an important step in the aggregation pathway.

Nonetheless, as we show next, the conformation assumed in the fibril appears to be encoded by the primary sequence. This encoding is hinted at by the results in Fig. 5a and b. Figure 7a repeats intrinsic β-propensities for $A\beta_{1-42}$ at 310 K and juxtaposes them to a cartoon representation of Riek *et al.*'s solution NMR model of $A\beta_{1-42}$ fibrils (Fig. 7b).[22] In

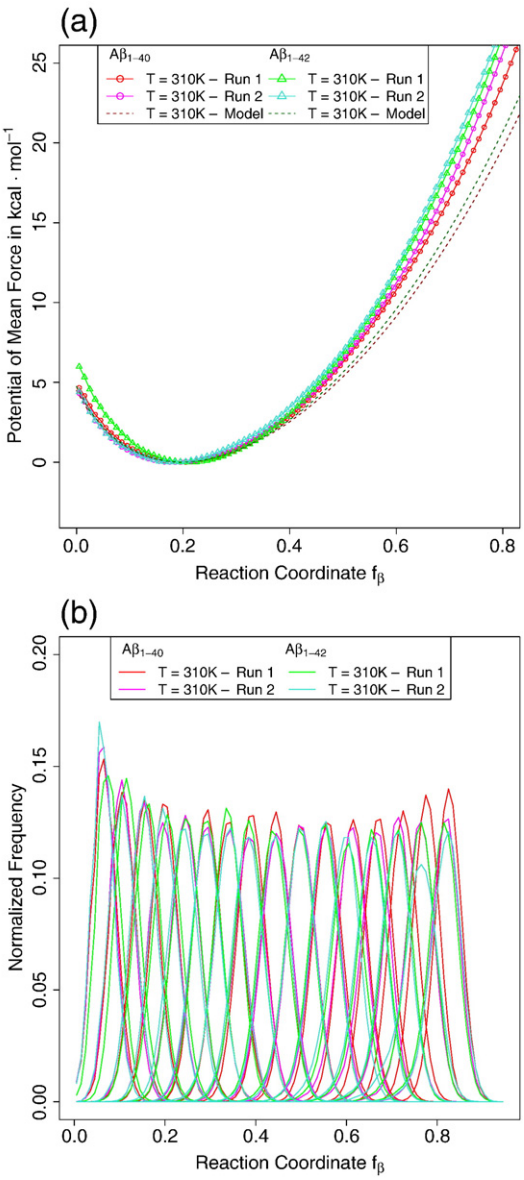
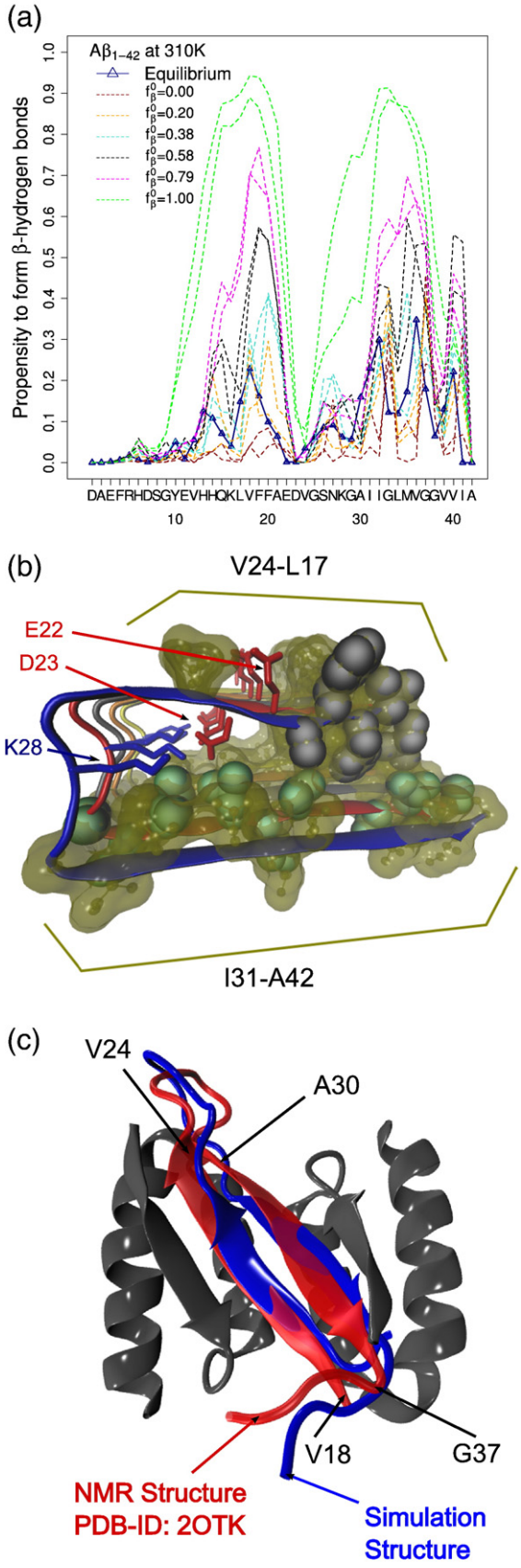


Fig. 6. PMF and umbrella sampling statistics as a function of a reaction coordinate reflecting net β-content. (a) PMF. Using umbrella sampling on the reaction coordinate f_β (see Materials and Methods), we obtained the PMF for both alloforms. We performed two massive REMC runs and show their results at 310 K individually, as there is no way to obtain a clean standard error. Differences between runs appear to be comparable to those between peptides, indicating that $A\beta_{1-40}$ and $A\beta_{1-42}$ are not significantly different in this analysis. Broken lines without symbols show synthetic data created by assuming a binomial distribution with a probability of 0.2 (see text). (b) Histograms of f_β for restrained runs at 310 K for all 16 equilibrium positions f_β^0. These data demonstrate excellent overlap between neighboring windows and lend confidence to the results in (a).



addition, we plot in Fig. 7a the propensities to form β-hydrogen bonds observed in the restrained (i.e., biased) simulations at a few roughly evenly spaced values of f_β^0. It becomes obvious that there is excellent agreement between the locations of the β-strands in the fibril and the sequence-encoded β-propensity. Excluding the biased simulations at $f_\beta^0 = 0.0$ and $f_\beta^0 = 1.0$, we find that the most populated regions are universally the CHC (L17-A21) and the C-terminal region starting at I31. We note that the C-terminal glycine residues act as structure breakers, and that this behavior may only be overridden in the presence of very strong ordering restraints. We further note that the equilibrium results in Fig. 7a do in fact most closely resemble the restrained results with $f_\beta^0 = 0.2$, which can be expected given the minimum position of the PMF at $f_\beta = 0.2$, as seen in Fig. 6a. This mutual consistency is highly encouraging for either set of results given that they are obtained from completely independent Monte Carlo runs. Lastly, we wish to emphasize that f_β and the propensity to form β-hydrogen bonds are conceptually independent, albeit weakly correlated, parameters.[38] The consistency of results in Fig. 7a is therefore not at all encoded by our methodology.

Fig. 7. Comparison of Aβ's intrinsic tendency to form β-hydrogen bonds in an ordering environment (a) to experimentally determined structures (b and c).[22,57] (a) Normalized frequencies for $A\beta_{1–42}$ akin to Fig. 6b for the equilibrium data at 310 K, as well as for both restrained (biased) runs at various values of the umbrella minimum position f_β^0 indicated in the figure legend. (b) A rendering of Riek *et al.*'s $A\beta_{1–42}$ fibril model (PDB ID 2BEG) with a cartoon representation of the backbone; transparent molecular surface and CPK representations of all hydrophobic side chains; and licorice representations of charged side chains. Additionally, residues L17-A21 (CHC) in the chain nearest to the viewer are highlighted in space-filling representation (gray), as are those of the C-terminal hydrophobic domain of the second chain (green), illustrating the characteristic registry. Note that in the fibril: (i) all β-hydrogen bonds are intermolecular; (ii) the salt bridge D23↔K28 was imposed as a constraint during structure determination; (iii) the model is actually obtained for $^{35Mox}A\beta_{1–42}$ fibrils; and (iv) residues 1–16 could not be resolved.[22] Also note that very similar results are obtained for the $A\beta_{1–40}$ fibril models of Petkova *et al.*[21] (c) $A\beta_{1–40}$ bound to an affibody protein identified via phage display (PDB ID 2OTK).[57] The protein is shown in cartoon representation in gray, along with the experimentally determined structure of $A\beta_{1–40}$ (transparent red). Positions of specific residues are shown via arrows and labels. Restrained simulation data at the four highest values of f_β^0 were combined and clustered (see Materials and Methods). The simulation structure (blue) shown is a member of a cluster representing 3.4% of the ensemble considered. Note that only experimentally resolved residues 16–40 are shown. Rendering was performed using VMD.[47]

As an additional comparison, Fig. 7c shows an alignment of a simulation structure taken from restrained simulations of $A\beta_{1-40}$ with large values of f_β^0 to an experimentally determined "monomeric" structure of $A\beta_{1-40}$ bound to an "affibody" dimer (i.e., an artificial mutant protein identified via phage display).[57] We show this qualitative comparison to make two points: (1) even a binding pocket enclosing Aβ almost completely resembles an ordering "field" that is representable in good approximation by the umbrella potential on f_β, and (2) the sequence-specific tendency to form β-hydrogen bonds in particular regions of the peptide (compare Fig. 5a) seems to be conserved between the fibril and the protein-bound structure in Fig. 7. The latter point reinforces that these are largely sequence-intrinsic properties.

Discussion and Conclusions

We have shown quantitative evidence that monomeric Aβ adopts a micelle-like architecture in aqueous solution. We quantified collapse (Fig. 1a), average topology (Figs. 1b, 2, 3, and 4b), disorder (Figs. 3, 4a, and 5), and intrinsic β-propensity (Figs. 5–7). As discussed both below and above, while some of these results are supported explicitly and/or implicitly by different experimental and simulation data, the emerging statistically significant atomic-resolution picture constitutes a novel result. In detail, we can extract from our results a technical observation and four main conclusions.

Simple restraint potentials represent powerful tools to unmask intrinsic sequence specificity

The results in Fig. 7 show that the conformation of the monomer determined experimentally in the presence of either the fibril or a suitable binding partner shares features with an ensemble generated in the presence of an ordering bias favoring the β-region of ϕ/ψ space. This suggests that it may be possible to coarsely mimic complex solution conditions with simple potentials. This powerful methodology may allow us in the future to define a conformational inventory for Aβ, from which to construct atomic-resolution pathways of supramolecular assembly by using the said inventory as a "basis set."

Monomeric Aβ populates a collapsed disordered ensemble

NMR studies have suggested that the hydrodynamic properties of Aβ resemble those of chemically denatured proteins (i.e., that the chains are elongated and swollen in aqueous solution).[58] This is based on a combined scaling analysis of Aβ fragments of varying lengths together with denatured proteins and very short peptides and amino acids. However, a large body of work, including the data presented in Figs. 1–4, contradict this finding. In fact, Danielsson *et al.* themselves pointed out (i) that the scaling exponent when only considering the Aβ-derived fragments is 0.44, and not 0.5; (ii) that this indicates the onset of a transition from elongated shorter fragments to collapsed longer ones; and (iii) that hydrodynamic radii are poor reporters of size for nonspherical objects. Conversely, most simulation studies,[15,32–34] the structure of Zhang *et al.*,[42] and the study of Chen and Glabe[59] show or suggest that the full-length alloforms are mostly collapsed (compare Figs. 1 and 3). In fact, considerations of finite-size effects cast doubts on whether a scaling law constructed from fragments of Aβ can be representative of the ensemble populated by the full-length alloforms. Other NMR studies may interpret the apparent absence of medium-range and long-range nuclear Overhauser enhancements as indicative of extended structures,[10] ignoring the fact that globules with a highly fluid core produce no such nuclear Overhauser enhancements.[60] It is nonetheless worth clarifying that Aβ is a disordered heteropolymer in which local behavior will not uniformly mirror global behavior. For example, consistent with the data presented here, the N-terminus of Aβ will often assume a locally extended and dissociated conformation that contrasts with the behavior of the rest of the peptide. This is apparent from Figs. 2 and 3 in particular.

Aβ's intrinsic disorder has been established primarily by solution NMR[8,10] and CD.[53] The interesting structure of $A\beta_{10-35}$ obtained by Zhang *et al.*[42] has been tackled by simulation work pointing out that a predominantly disordered ensemble is capable of satisfying the same NMR restraints.[45,46] The total secondary structure contents along the simulations (Fig. 5) are in good agreement with the CD estimates of Fezoui and Teplow, who predicted, via spectral deconvolution, a 5% α-content and a 25% β-content in aqueous buffer at 298 K.[53]

Oligomerization and aggregation of Aβ depend on hydrophobic interactions

Through analysis of temperature-dependent aggregation rates, Lin *et al.* proposed the nucleation of Aβ aggregation to be dependent on hydrophobic interactions.[56] Similarly, a systematic analysis of mutants found that charge content and hydrophobicity are the dominant modifiers of aggregation propensity.[61,62] Our data offer the suggestion that, for conversion into a fibrillar aggregate, intramolecular hydrophobic interactions (Fig. 2) have to be replaced by intermolecular ones. Only the presence of a template or oligomer "phase"

would facilitate this process enough for it to occur spontaneously, giving rise to lag time kinetics. Specifically, individual conformations of monomeric Aβ may be described as possessing an imperfect micelle-like architecture characterized by fluctuating patches of hydrophobic surface, as indicated by Fig. 4c and d. These data also point out an interesting difference between the two alloforms. It seems as if C-terminal hydrophobic residues in $A\beta_{1-42}$ are sequestered more efficiently from solvent. This intuitive trend stands in contrast to the less efficient sequestration of residues L17-A21 (higher solvent accessibilities) in $A\beta_{1-42}$, which, unfortunately, we have to treat with caution due to concerns about statistical significance. This nonetheless provides links to several pieces of experimental data:

(1) The less efficient protection of the CHC may provide an intuitive explanation for the universally faster aggregation kinetics of $A\beta_{1-42}$ irrespective of mechanistic details.
(2) The less solvent-accessible C-terminus agrees with suggestions from the literature concerning its "rigidity."[63] Like others,[33,63] we see increased β-content in the C-terminal hydrophobic region (Fig. 5), but would hesitate to assign to this finding relevance to the differential aggregation properties of $A\beta_{1-40}$ and $A\beta_{1-42}$.
(3) It is not difficult to imagine that the specificity for forming small oligomers of certain sizes is dependent on the imperfect micelle-like architecture of the monomer (Figs. 3 and 4b) given that large-scale structural changes and full entanglement for such relatively short peptides would not appear feasible. It is therefore also reasonable to conclude that differences in oligomer distributions between the two alloforms[13] can be explained by differences in the micelle-like architecture between strictly monomeric $A\beta_{1-40}$ and $A\beta_{1-42}$. To avoid confusion, it should be noted that oligomers of Aβ have themselves been described as "micelles."[16]
(4) Specifically, we may conjecture that the spherical nonfibrillar micellar oligomers suggested by Lomakin *et al.* will be capped in size by successive sequestration of more and more accessible hydrophobic surface areas.[16] This agrees well with the suggestions of Fawzi *et al.* on the trends of hydrophobic density with increasing oligomer size for protofibrillar aggregates.[64] Through quaternary arrangements of amyloid fibrils (not visible in Fig. 7b), hydrophobic residues can potentially be sequestered with particular efficiency, as observed in simulations of a coarse-grained amphipathic peptide.[65]

Fibril elongation occurs via templated assembly

It appears likely that fibril elongation via monomer addition follows a templated assembly mechanism,[66,67] as the β-sheet propensity of the monomer is weak (Figs. 5 and 6). Presumably, the first contact ("docking" step) is mediated by nonspecific hydrophobic interactions similar to what is outlined in the previous paragraph. The "locking" step then consists of a conformational transition of the docked monomer into that observed in the fibril.

If we introduce an ordering potential favoring extended states of the polypeptide backbone, the monomer, on its own, populates conformations compatible with those in the fibril model and bound to an affibody (Fig. 7). Here, the latter represents a different type of template. This result, in conjunction with the observed sequence-specific β-propensities at equilibrium (Fig. 5), suggests that the primary sequence does inform the template-assisted structural conversion (i.e., the locking step). However, based on our data, the overlap between the equilibrium monomer ensemble and ordered structures, as shown in Fig. 7, is too small to support the hypothesis that elongation would occur via conformational selection.[68] In such a model, only monomers in fibril-competent conformation would successfully dock with no requirement for a subsequent locking step.

The small overlap between monomer ensemble and conformations within ordered assemblies is also more consistent with recently corroborated views on polymorphism in Aβ fibrils.[69–71] Electron cryomicroscopy data suggest that many different fibril morphologies can be obtained even under identical solution conditions.[70] This would indicate substantial heterogeneity in terms of the quaternary, tertiary, and even sequence-specific details of secondary structure. Our result—that monomeric Aβ is dominantly disordered—supports the hypothesis that morphologically different templates may be nucleated initially, and that the template will control monomer conformation upon elongation. The observed preferences under bias shown in Fig. 7a are coarse, and Fig. 7b is in no way meant to indicate that the ordering potential predisposes Aβ to a unique monomer conformation consistent with only a single structure of amyloid fibril. Templated assembly was very recently demonstrated experimentally.[71] Fibrils with differing morphologies were obtained and used in elongation experiments under identical solution conditions. It could be shown that most elongated fibrils retain the biophysical characteristics of their parent seeds. This is a particularly clear demonstration of the "polyvalence" of disaggregated Aβ to participate in different supramolecular assemblies that share only coarse structural hallmarks, and is also consistent with our results as outlined above.

E22 and D23 are topologically unique and may control fibril architecture

The majority of known mutants in the human amyloid precursor protein are not located within $A\beta_{1-43}$. Of those that are and have been characterized, a majority affect position E22. *In vivo* data will be influenced by multitudinous cellular factors, including expression levels or production rates via proteolysis. These factors may well account for the vast majority of impact a given mutation has. Moreover, many possible mutations may not yield viable cells, which may bias our perception.

Nonetheless, it seems surprising that residues 22 and 23 appear to play a unique role in the conformational ensembles studied here as well. As Fig. 7b shows, E22, D23, and K28 are the only ionic residues within the ordered region of the fibril (i.e., one anionic group will have to be compensated for by the electrolyte bath, while the remaining two residues form a salt bridge).[72] The resultant linear charge density represents an important feature of the fibril. It poses intuitive explanations of the result that aggregation rates seem to increase with lower pH and higher ionic strength.[56] Our data show no evidence for the formation of a salt bridge between K28 and either D23 or E22 (see Fig. 2) at the monomer level. This is indirectly supported by the result that chemical cross-linking has a huge impact on aggregation kinetics,[73] which would not be expected had the salt bridge been preformed already. What we do see is that the requirement to keep both anionic residues solvent accessible poses a topological constraint (Fig. 3).

It is imaginable that such a constraint controls the details of the micelle-like architecture described above for monomeric Aβ, and that nonconservative mutations in those positions will therefore have a profound and nontrivial impact on oligomerization and aggregation mechanism from the earliest steps onward. We plan to study several mutants by a similar approach in the future, which may enable us to rationalize experimental results such as those of Päiviö *et al.* and to ultimately offer predictions about the aggregation characteristics of given mutants.[74]

Materials and Methods

Equilibrium Monte Carlo simulations of the full-length alloforms of Aβ

Uncapped $A\beta_{1-40}$ or $A\beta_{1-42}$ with charged termini at all-atom resolution was placed in a spherical simulation droplet of 65 Å radius in completely randomized conformations. Explicit sodium and chloride ions were added to neutralize the peptides' charge and to add a background electrolyte concentration of ~130 mM. Glutamate, aspartate, lysine, and arginine side chains were all modeled as charged, while histidines were considered neutral. These choices mimic neutral pH conditions. Protonation states are an important consideration for all simulation studies of Aβ, and results are expected to vary significantly with the assumed simulation pH.[75] All particles were restrained to remain inside the spherical simulation volume by a boundary potential equivalent to previous work.[39]

Monte Carlo simulations in rigid-body/torsional space were performed for 10^8 elementary steps in the NVT ensemble at four different temperatures spaced evenly from 280 K to 325 K. For enhancement of sampling, each run employed the replica-exchange technique[76] to attempt structure exchanges between neighboring temperatures every 10^4 steps. Twelve runs were performed for each peptide. The exploration of conformational space in Monte Carlo simulations depends crucially on the moveset. Largely in accordance with prior work,[39] we used a complex moveset consisting of approximately 10% rigid-body moves, 14% side-chain moves, 6% ω moves, and 55% φ/ψ moves. Here we augmented the moveset by using exact concerted rotation moves (in φ/ψ space only) combining the closure strategy of Dinner[77] with the biased prerotation idea of Favrin *et al.*[78] These moves made up the remaining 15% of the moveset. We verified that the employed moveset is unbiased and ergodic (data not shown) within the selected degrees of freedom.[36]

Aqueous solvation effects were described by the ABSINTH implicit solvation model. This includes the published choices for all solvation and interaction parameters.[36] We used 14-Å cutoffs for electrostatic interactions and 10-Å cutoffs for steric and dispersive terms. We included long-range electrostatic corrections computing interactions between charged groups in a monopole–monopole approximation if their distance exceeded the cutoff. All the methods described here are implemented in the CAMPARI software package, which is freely available for download‡.

Umbrella sampling runs to obtain the PMF

In prior work, Vitalis *et al.* studied the β-propensity of polyglutamine by biased simulations employing the reaction coordinate that measures the β-content of a polypeptide chain based on dihedral criteria alone.[38] Here, we repeated a very similar protocol by considering 16 different umbrella positions at f_β^0 values of 0.0, 0.07, 0.14, 0.2, 0.26, 0.32, 0.38, 0.44, 0.51, 0.58, 0.65, 0.72, 0.79, 0.86, 0.93, and 1.0, with a uniform k_β of 250.0 kcal/mol where the umbrella potential is $U_\beta = k_\beta(f_\beta - f_\beta^0)^2$. In addition, we chose four temperatures of a narrow range from 300 K to 330 K. The resultant 64 replicas were arranged in sequence such that, between neighbors, either the temperature was different by 10 K or the f_β^0 value changed to a nearest neighbor, but not both. This sequence defined the input for a two-dimensional REMC calculation of 1.5×10^8 steps in total. Swaps between neighboring replicas were attempted every 2×10^4 steps. For each of the two peptides, we performed two completely independent runs.

‡ http://sourceforge.net/projects/campari

Histograms of f_β were collected every 20 steps. They were used to construct the PMF along f_β via the weighted histogram analysis method.[79,80] Both the convergence of the free-energy estimates between windows and the coverage of the possible f_β values suggested excellent sampling quality.

Analysis of data and construction of error estimates

Most analyses performed in this article, including the DSSP assignments originally developed by Kabsch and Sander,[54] are built into the CAMPARI software package. Internal scaling and angular correlation analyses are defined exactly as in prior work.[81] The solvent-accessible volume (SAV) fractions η are intrinsic parameters of the ABSINTH model and are explained in detail elsewhere.[36] SAV data were collected every 2×10^5 steps. Pairwise RMSD values were computed by VMD[47] at the frequency indicated in the caption to Fig. 4. All other quantities were collected every 10^3 steps. For the equilibrium simulations, the first 10^7 steps were discarded as equilibration, while for the restrained simulations we chose a larger interval of 3×10^7 steps due to the more rugged energy landscape created by the additional term to the Hamiltonian.

The clustering employed to obtain Fig. 3 was performed using Wordom.[82] We used a leader-like algorithm on the 280-K ensembles and considered the C^α and C^β atoms of residues 12–40, as well as the C^ζ atoms of F19 and F20, as the alignment set. This was necessary to be able to obtain clusters with meaningful sizes even though in RMSD space the effective cluster radius that we used was still large (4.0 Å). For $A\beta_{1-40}$ ($A\beta_{1-42}$), we obtained 748 (494) clusters, with the largest one encompassing 3525 (5692) out of 108,000 snapshots. For the analysis in Fig. 7c, we followed a similar protocol. Data from both restrained runs of $A\beta_{1-40}$ for the four highest f_β^0 values were combined and clustered using the three main-chain backbone atoms of residues 16–40 as the alignment criterion. We considered only every 10th of the 9.6×10^5 snapshots here. This backbone-centric analysis yielded 593 clusters, the largest of which encompassed 8869 of 96,000 structures. The 10 most populated clusters were analyzed to obtain pairwise RMSD values for the first model of $A\beta_{1-40}$ in the NMR ensemble of an affibody protein bound to $A\beta_{1-40}$ (PDB ID 2OTK).[57] We found a cluster encompassing 3293 structures with RMSD values in the 2.5–5 Å regime and selected the member with the lowest RMSD distance to the NMR structure for display in Fig. 7c. It is likely that we would have found even more similar structures by considering all data from the restrained simulations. Note that this clustering of subensembles of restrained simulations is a purely qualitative analysis and provides no thermodynamic insight.

Standard errors were estimated by a modified block averaging procedure[81] in which each individual trajectory (i.e., each temperature in each run for each peptide) is considered to be a single block. In this way, for a given temperature, there are 12 blocks from which we constructed standard errors as the standard deviation between trajectory averages. This procedure mimics performing truly independent replicas of the same experiment, as all interblock correlations are, by design, eliminated. Minimum/maximum error estimates would have generally allowed similar conclusions, although the size of the errors would have been somewhat larger as expected.

Acknowledgements

This work was supported by the Swiss National Science Foundation (by a grant to A.C.) and the Forschungskredit of the University of Zurich (by a grant to A.V.). Most calculations were carried out on the Schrödinger computing cluster at the University of Zurich.

References

1. Hardy, J. A. & Higgins, G. A. (1992). Alzheimer's disease: the amyloid cascade hypothesis. *Science*, **256**, 184–185.
2. Dahlgren, K. N., Manelli, A. M., Stine, W. B., Baker, L. K., Krafft, G. A. & LaDu, M. J. (2002). Oligomeric and fibrillar species of amyloid-β peptides differentially affect neuronal viability. *J. Biol. Chem.* **277**, 32046–32053.
3. Jarrett, J. T., Berger, E. P. & Lansbury, P. T. (1993). The carboxy terminus of the β amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry*, **32**, 4693–4697.
4. Frieden, C. (2007). Protein aggregation processes: in search of the mechanism. *Protein Sci.* **16**, 2334–2344.
5. Eliezer, D. (2009). Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **19**, 23–30.
6. Zagorski, M. G., Yang, J., Shao, H., Ma, K., Zeng, H. & Hong, A. (1999). Methodological and chemical factors affecting amyloid-β peptide amyloidogenicity. *Methods Enzymol.* **309**, 189–204.
7. Etienne, M. A., Edwin, N. J., Aucoin, J. P., Russo, P. S., McCarley, R. L. & Hammer, R. P. (2007). β-Amyloid protein aggregation. in: *Peptide Characterization and Application Protocols*, pp. 203–225, Humana Press, Clifton, NJ.
8. Riek, R., Guntert, P., Döbeli, H., Wipf, B. & Wüthrich, K. (2001). NMR studies in aqueous solution fail to identify significant conformational differences between the monomeric forms of two Alzheimer peptides with widely different plaque-competence, Aβ(1–40)(ox) and Aβ(1–42)(ox). *Eur. J. Biochem.* **268**, 5930–5936.
9. Danielsson, J., Jarvet, J., Damberg, P. & Gräslund, A. (2005). The Alzheimer β-peptide shows temperature-dependent transitions between left-handed 3(1)-helix, β-strand and random coil secondary structures. *FEBS J.* **272**, 3938–3949.
10. Hou, L. M., Shao, H. Y., Zhang, Y. B., Li, H., Menon, N. K., Neuhaus, E. B. *et al.* (2004). Solution NMR studies of the Aβ(1–40) and Aβ(1–42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *J. Am. Chem. Soc.* **126**, 1992–2005.
11. Sticht, H., Bayer, P., Willbold, D., Dames, S., Hilbich, C., Bayreuther, K. *et al.* (2004). Structure of amyloid

A4-(1–40)-peptide of Alzheimer's disease. *Eur. J. Biochem.* **233**, 293–298.
12. Watson, A. A., Fairlie, D. P. & Craik, D. J. (1998). Solution structure of methionine-oxidized amyloid β-peptide (1–40). Does oxidation affect conformational switching? *Biochemistry*, **37**, 12700–12706.
13. Bitan, G., Kirkitadze, M. D., Lomakin, A., Vollers, S. S., Benedek, G. B. & Teplow, D. B. (2003). Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proc. Natl Acad. Sci. USA*, **100**, 330–335.
14. Benseny-Cases, N., Cocera, M. & Cladera, J. (2007). Conversion of non-fibrillar β-sheet oligomers into amyloid fibrils in Alzheimer's disease amyloid peptide aggregation. *Biochem. Biophys. Res. Commun.* **361**, 916–921.
15. Bernstein, S. L., Wyttenbach, T., Baumketner, A., Shea, J. E., Bitan, G., Teplow, D. B. *et al.* (2005). Amyloid β-protein: monomer structure and early aggregation states of Aβ 42 and its Pro(19) alloform. *J. Am. Chem. Soc.* **127**, 2075–2084.
16. Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A. & Teplow, D. B. (1996). On the nucleation and growth of amyloid β-protein fibrils: detection of nuclei and quantitation of rate constants. *Proc. Natl Acad. Sci. USA*, **93**, 1125–1129.
17. Kirkitadze, M. D., Condron, M. M. & Teplow, D. B. (2001). Identification and characterization of key kinetic intermediates in amyloid β-protein fibrillogenesis. *J. Mol. Biol.* **312**, 1103–1119.
18. Sabate, R., Gallardo, M. & Estelrich, J. (2005). Temperature dependence of the nucleation constant rate in β amyloid fibrillogenesis. *Int. J. Biol. Macromol.* **35**, 9–13.
19. Fändrich, M. (2007). Absolute correlation between lag time and growth rate in the spontaneous formation of several amyloid-like aggregates and fibrils. *J. Mol. Biol.* **365**, 1266–1270.
20. Carrotta, R., Manno, M., Bulone, D., Martorana, V. & San Biagio, P. L. (2005). Protofibril formation of amyloid β-protein at low pH via a non-cooperative elongation mechanism. *J. Biol. Chem.* **280**, 30001–30008.
21. Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F. *et al.* (2002). A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl Acad. Sci. USA*, **99**, 16742–16747.
22. Lührs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Döbeli, H. *et al.* (2005). 3D structure of Alzheimer's amyloid-β(1–42) fibrils. *Proc. Natl Acad. Sci. USA*, **102**, 17342–17347.
23. Fändrich, M., Meinhardt, J. & Grigorieff, N. (2009). Structural polymorphism of Alzheimer Aβ and other amyloid fibrils. *Prion*, **3**, 89–93.
24. Paravastu, A. K., Leapman, R. D., Yau, W. M. & Tycko, R. (2008). Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl Acad. Sci. USA*, **105**, 18349–18354.
25. Miller, Y., Ma, B. & Nussinov, R. (2010). Polymorphism in Alzheimer Aβ amyloid organization reflects conformational selection in a rugged energy landscape. *Chem. Rev.* **110**, 4820–4838.
26. Walsh, D. M. & Selkoe, D. J. (2007). Aβ oligomers—a decade of discovery. *J. Neurochem.* **101**, 1172–1184.
27. Ono, K., Condron, M. M. & Teplow, D. B. (2009). Structure–neurotoxicity relationships of amyloid β-protein oligomers. *Proc. Natl Acad. Sci. USA*, **106**, 14745–14750.
28. Elcock, A. (2010). Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol.* **20**, 196–206.
29. Urbanc, B., Betnel, M., Cruz, L., Bitan, G. & Teplow, D. B. (2010). Elucidation of amyloid β-protein oligomerization mechanisms: discrete molecular dynamics study. *J. Am. Chem. Soc.* **132**, 4266–4280.
30. Horn, A. H. C. & Sticht, H. (2010). Amyloid-β42 oligomer structures from fibrils: a systematic molecular dynamics study. *J. Phys. Chem. B*, **114**, 2219–2226.
31. Zheng, J., Yu, X., Wang, J. D., Yang, J. C. & Wang, Q. M. (2010). Molecular modeling of two distinct triangular oligomers in amyloid β-protein. *J. Phys. Chem. B*, **114**, 463–470.
32. Yang, M. & Teplow, D. B. (2008). Amyloid β-protein monomer folding: free-energy surfaces reveal alloform-specific differences. *J. Mol. Biol.* **384**, 450–464.
33. Sgourakis, N. G., Yan, Y., Mccallum, S. A., Wang, C. & Garcia, A. E. (2007). The Alzheimer's peptides Aβ40 and 42 adopt distinct conformations in water: a combined MD/NMR study. *J. Mol. Biol.* **368**, 1448–1457.
34. Li, W. F., Zhang, J., Su, Y., Wang, J., Qin, M. & Wang, W. (2007). Effects of zinc binding on the conformational distribution of the amyloid-β peptide based on molecular dynamics simulations. *J. Phys. Chem. B*, **111**, 13814–13821.
35. Xu, Y. C., Shen, J. J., Luo, X. M., Zhu, W. L., Chen, K. X., Ma, J. P. *et al.* (2005). Conformational transition of amyloid β-peptide. *Proc. Natl Acad. Sci. USA*, **102**, 5403–5407.
36. Vitalis, A. & Pappu, R. V. (2009). ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J. Comput. Chem.* **30**, 673–699.
37. Vitalis, A., Wang, X. & Pappu, R. V. (2008). Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *J. Mol. Biol.* **384**, 279–297.
38. Vitalis, A., Lyle, N. & Pappu, R. V. (2009). Thermodynamics of β-sheet formation in polyglutamine. *Biophys. J.* **97**, 303–311.
39. Williamson, T. E., Vitalis, A., Crick, S. L. & Pappu, R. V. (2010). Modulation of polyglutamine conformations and dimer formation by the N-terminus of Huntingtin. *J. Mol. Biol.* **396**, 1295–1309.
40. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. (2010). Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl Acad. Sci. USA*, **107**, 8183–8188.
41. Zhang, S., Casey, N. & Lee, J. P. (1998). Residual structure in the Alzheimer's disease peptide: probing the origin of a central hydrophobic cluster. *Fold. Des.* **3**, 413–422.
42. Zhang, S., Iwata, K., Lachenmann, M. J., Peng, J. W., Li, S., Stimson, E. R. *et al.* (2000). The Alzheimer's peptide Aβ adopts a collapsed coil structure in water. *J. Struct. Biol.* **130**, 130–141.

43. Shivaprasad, S. & Wetzel, R. (2006). Scanning cysteine mutagenesis analysis of Aβ-(1–40) amyloid fibrils. *J. Biol. Chem.* **281**, 993–1000.
44. Olofsson, A., Sauer-Eriksson, A. E. & Öhman, A. (2006). The solvent protection of Alzheimer amyloid-β-(1–42) fibrils as determined by solution NMR spectroscopy. *J. Biol. Chem.* **281**, 477–483.
45. Baumketner, A. & Shea, J. E. (2007). The structure of the Alzheimer amyloid β 10–35 peptide probed through replica-exchange molecular dynamics simulations in explicit solvent. *J. Mol. Biol.* **366**, 275–285.
46. Ikebe, J., Kamiya, N., Ito, J. I., Shindo, H. & Higo, J. (2007). Simulation study on the disordered state of an Alzheimer's β amyloid peptide Aβ(12–36) in water consisting of random-structural, β-structural, and helical clusters. *Protein Sci.* **16**, 1596–1608.
47. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD —Visual Molecular Dynamics. *J. Mol. Graphics*, **14**, 33–38.
48. Schäfer, L. (1999). *Excluded Volume Effects in Polymer Solutions as Explained by the Renormalization Group.* Springer, Berlin, Germany.
49. Danielsson, J., Andersson, A., Jarvet, J. & Gräslund, A. (2006). N-15 relaxation study of the amyloid β-peptide: structural propensities and persistence length. *Magn. Reson. Chem.* **44**, S114–S121.
50. Cifra, P., Benková, Z. & Bleha, T. (2008). Persistence lengths and structure factors of wormlike polymers under confinement. *J. Phys. Chem. B*, **112**, 1367–1375.
51. Tjernberg, L. O., Näslund, J., Lindqvist, F., Johansson, J., Karlström, A. R., Thyberg, J. *et al.* (1996). Arrest of β-amyloid fibril formation by a pentapeptide ligand. *J. Biol. Chem.* **271**, 8545–8548.
52. Cecchini, M., Curcio, R., Pappalardo, M., Melki, R. & Caflisch, A. (2006). A molecular dynamics approach to the structural characterization of amyloid aggregation. *J. Mol. Biol.* **357**, 1306–1321.
53. Fezoui, Y. & Teplow, D. B. (2002). Kinetic studies of amyloid β-protein fibril assembly. *J. Biol. Chem.* **277**, 36948–36954.
54. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
55. Poulsen, S. A., Watson, A. A., Fairlie, D. P. & Craik, D. J. (2000). Solution structures in aqueous SDS micelles of two amyloid β peptides of Aβ(1–28) mutated at the α-secretase cleavage site (K16E, K16F). *J. Struct. Biol.* **130**, 142–152.
56. Lin, M. S., Chen, L. Y., Tsai, H. T., Wang, S. S. S., Chang, Y., Higuchi, A. *et al.* (2008). Investigation of the mechanism of β-amyloid fibril formation by kinetic and thermodynamic analyses. *Langmuir*, **24**, 5802–5808.
57. Hoyer, W., Grönwall, C., Jonsson, A., Ståhl, S. & Härd, T. (2008). Stabilization of a β-hairpin in monomeric Alzheimer's amyloid-β peptide inhibits amyloid formation. *Proc. Natl Acad. Sci. USA*, **105**, 5099–5104.
58. Danielsson, J., Jarvet, J., Damberg, P. & Gräslund, A. (2002). Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer Aβ(1–40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. *Magn. Reson. Chem.* **40**, S89–S97.
59. Chen, Y. & Glabe, C. G. (2006). Distinct early folding and aggregation properties of Alzheimer amyloid-β peptides Aβ40 and Aβ42. *J. Biol. Chem.* **281**, 24414–24422.
60. Ramboarina, S. & Redfield, C. (2003). Structural characterisation of the human α-lactalbumin molten globule at high temperature. *J. Mol. Biol.* **330**, 1177–1188.
61. Meinhardt, J., Tartaglia, G. G., Pawar, A., Christopeit, T., Hortschansky, P., Schroeckh, V. *et al.* (2007). Similarities in the thermodynamics and kinetics of aggregation of disease-related Aβ(1–40) peptides. *Protein Sci.* **16**, 1214–1222.
62. Tartaglia, G. G., Cavalli, A., Pellarin, R. & Caflisch, A. (2004). The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci.* **13**, 1939–1941.
63. Yan, Y. & Wang, C. (2006). Aβ42 is more rigid than Aβ40 at the C-terminus: implications for Aβ aggregation and toxicity. *J. Mol. Biol.* **364**, 853–862.
64. Fawzi, N. L., Okabe, Y., Yap, E. H. & Head-Gordon, T. (2007). Determining the critical nucleus and mechanism of fibril elongation of the Alzheimer's Aβ(1–40) peptide. *J. Mol. Biol.* **365**, 535–550.
65. Pellarin, R. & Caflisch, A. (2006). Interpreting the aggregation kinetics of amyloid peptides. *J. Mol. Biol.* **360**, 882–892.
66. Esler, W. P., Stimson, E. R., Jennings, J. M., Vinters, H. V., Ghilardi, J. R., Lee, J. P. *et al.* (2000). Alzheimer's disease amyloid propagation by a template-dependent dock-lock mechanism. *Biochemistry*, **39**, 6288–6295.
67. Pellarin, R., Guarnera, E. & Caflisch, A. (2007). Pathways and intermediates of amyloid fibril formation. *J. Mol. Biol.* **374**, 917–924.
68. Tsai, C. J., Ma, B., Sham, Y. Y., Kumar, S. & Nussinov, R. (2001). Structured disorder and conformational selection. *Proteins Struct. Funct. Genet.* **44**, 418–427.
69. Goldsbury, C. S., Wirtz, S., Müller, S. A., Sunderji, S., Wicki, P., Aebi, U. *et al.* (2000). Studies on the *in vitro* assembly of Aβ 1–40: implications for the search for Aβ fibril formation inhibitors. *J. Struct. Biol.* **130**, 217–231.
70. Meinhardt, J., Sachse, C., Hortschansky, P., Grigorieff, N. & Fändrich, M. (2009). Aβ(1–40) fibril polymorphism implies diverse interaction patterns in amyloid fibrils. *J. Mol. Biol.* **386**, 869–877.
71. Kodali, R., Williams, A. D., Chemuru, S. & Wetzel, R. (2010). Aβ(1–40) forms five distinct amyloid structures whose β-sheet contents and fibril stabilities are correlated. *J. Mol. Biol.* **401**, 503–517.
72. Petkova, A. T., Yau, W. M. & Tycko, R. (2006). Experimental constraints on quaternary structure in Alzheimer's β-amyloid fibrils. *Biochemistry*, **45**, 498–512.
73. Sciarretta, K. L., Gordon, D. J., Petkova, A. T., Tycko, R. & Meredith, S. C. (2005). Aβ40-lactam(D23/K28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry*, **44**, 6003–6014.
74. Päiviö, A., Jarvet, J., Gräslund, A., Lannfelt, L. & Westlind-Danielsson, A. (2004). Unique physicochemical profile of β-amyloid peptide variant Aβ 1–40E22G protofibrils: conceivable neuropathogen in arctic mutant carriers. *J. Mol. Biol.* **339**, 145–159.

75. Khandogin, J. & Brooks, C. L. (2007). Linking folding with aggregation in Alzheimer's β-amyloid peptides. *Proc. Natl Acad. Sci. USA*, **104**, 16880–16885.
76. Swendsen, R. H. & Wang, J. S. (1986). Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **57**, 2607–2609.
77. Dinner, A. R. (2000). Local deformations of polymers with nonplanar rigid main-chain internal coordinates. *J. Comput. Chem.* **21**, 1132–1144.
78. Favrin, G., Irbäck, A. & Sjunnesson, F. (2001). Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. *J. Chem. Phys.* **114**, 8154–8158.
79. Kumar, S., Rosenberg, J. S., Bouzida, D., Swendsen, R. H. & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules: I. The method. *J. Comput. Chem.* **13**, 1011–1021.
80. Roux, B. (1995). The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **91**, 275–282.
81. Vitalis, A., Wang, X. L. & Pappu, R. V. (2007). Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys. J.* **93**, 1923–1937.
82. Seeber, M., Cecchini, M., Rao, F., Settanni, G. & Caflisch, A. (2007). Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*, **23**, 2625–2627.