In silico study of amyloid β-protein folding and oligomerization

B. Urbanc*†, L. Cruz*, S. Yun*, S. V. Buldyrev*‡, G. Bitan§, D. B. Teplow§, and H. E. Stanley*

*Center for Polymer Studies, Department of Physics, Boston University, Boston, MA 02215; and §Center for Neurologic Diseases, Brigham and Women's Hospital, and Department of Neurology, Harvard Medical School, Boston, MA 02115

Contributed by H. E. Stanley, November 2, 2004

Experimental findings suggest that oligomeric forms of the amyloid β protein (Aβ) play a critical role in Alzheimer's disease. Thus, elucidating their structure and the mechanisms of their formation is critical for developing therapeutic agents. We use discrete molecular dynamics simulations and a four-bead protein model to study oligomerization of two predominant alloforms, Aβ40 and Aβ42, at the atomic level. The four-bead model incorporates backbone hydrogen-bond interactions and amino acid-specific interactions mediated through hydrophobic and hydrophilic elements of the side chains. During the simulations we observe monomer folding and aggregation of monomers into oligomers of variable sizes. Aβ40 forms significantly more dimers than Aβ42, whereas pentamers are significantly more abundant in Aβ42 relative to Aβ40. Structure analysis reveals a turn centered at Gly-37–Gly-38 that is present in a folded Aβ42 monomer but not in a folded Aβ40 monomer and is associated with the first contacts that form during monomer folding. Our results suggest that this turn plays an important role in Aβ42 pentamer formation. Aβ pentamers have a globular structure comprising hydrophobic residues within the pentamer's core and hydrophilic N-terminal residues at the surface of the pentamer. The N termini of Aβ40 pentamers are more spatially restricted than Aβ42 pentamers. Aβ40 pentamers form a β-strand structure involving Ala-2–Phe-4, which is absent in Aβ42 pentamers. These structural differences imply a different degree of hydrophobic core exposure between pentamers of the two alloforms, with the hydrophobic core of the Aβ42 pentamer being more exposed and thus more prone to form larger oligomers.

Alzheimer's disease | discrete molecular dynamics | four-bead protein model | oligomer formation

The amyloid β-protein (Aβ) has been strongly linked to the etiology and pathogenesis of Alzheimer's disease (AD). Aβ assembles into amyloid fibrils and smaller, oligomeric assemblies. Experimental and clinical findings suggest that protofibrillar intermediates (1–3) and oligomeric forms (4–13) of Aβ may be particularly important. If so, elucidating the structures of these Aβ oligomers and the mechanisms of their formation is critical for developing therapeutic agents. Unlike proteins with stable folds, Aβ oligomers are metastable. They cannot be crystallized for x-ray diffraction studies nor can they be easily studied by using solution-phase NMR. Monomers and oligomers are also in dynamic equilibrium, which makes the study of pure populations of conformers using classical biophysical techniques difficult.

Aβ exists in two predominant forms, 40 (Aβ40) or 42 (Aβ42) amino acids in length. Of the two, Aβ42 is associated most strongly with an increased risk for AD, is more neurotoxic, and forms fibrils significantly faster. Recent experiments demonstrated that Aβ oligomers can be covalently cross-linked, and therefore stabilized, by using the technique of photo-induced cross-linking of unmodified proteins (PICUP) (14). During PICUP coupled with size-exclusion chromatography, Aβ40 and Aβ42 display distinct oligomer size distributions: Aβ40 displays a rapid equilibrium among monomers, dimers, trimers, and tetramers, whereas Aβ42 preferentially forms pentamer/hexamer units (paranuclei), which further assemble into beaded superstructures similar to early protofibrils (15). Additional studies of primary structure elements controlling early oligomerization demonstrate that Ile-41 is critical for paranucleus formation by Aβ42 and that Ala-42 is necessary for further assembly of Aβ42 into larger oligomers (16). In addition, oxidation of Met-35 blocks paranucleus formation but does not alter the Aβ40 oligomer size distribution (17).

Here we use *in silico* techniques to determine, at the atomic level, how Aβ monomers fold and assemble into oligomers. Traditional all-atom molecular dynamics (MD) with explicit solvent is an ideal method for studying Aβ oligomerization. However, even when using advanced technologies such as worldwide distributed computing, all-atom MD is limited to the study of aggregation processes occurring over time periods not exceeding ≈500 ns (18–20). *In vivo* and *in vitro* studies suggest that the time regime of Aβ oligomerization is measured in seconds to weeks (15, 21), at least 7 orders of magnitude greater than that accessible by all-atom MD. To overcome this temporal barrier and enable the study of Aβ folding and assembly, we combined an efficient discrete MD (DMD) algorithm with a coarse-grained protein model (22–32). This simulation approach produces oligomerization speeds $\approx 10^{10}$ greater than those obtainable with traditional MD. This increase in simulation speed allows us to simulate a relatively large number of peptides and thus obtain statistically significant results. We discuss here basic features of the four-bead *ab initio* DMD model and show how its use in simulating Aβ oligomerization produces important insights into Aβ alloform-specific folding and assembly events.

Methods

In our approach, we apply the DMD method, in which pairs of particles interact by means of spherically symmetric potentials consisting of one or more square wells (for an introduction to the method, see ref. 33). The DMD simulation method has been adapted successfully to model proteins (23, 25, 26, 30) and used to study folding and aggregation of a three-helix-bundle protein (22, 24), the SH3 protein (27–29), and Aβ (31, 32).

The Four-Bead Protein Model with Hydrogen-Bond Interactions. In our simulations we apply the four-bead protein model introduced by Ding *et al.* (30). In four-bead models (25, 26, 30, 34), each amino acid is replaced by at most four beads. These beads correspond to the amide N, the α-carbon C_α, and the carbonyl C′ groups. The fourth bead, representing the amino acid side-chain groups of atoms, is placed at the center of the nominal C_β atom. Because of their lack of side chains, glycines are represented by only three beads. A full description of the four-bead protein model imple-

Abbreviations: Aβ, amyloid β-protein; AD, Alzheimer's disease; MD, molecular dynamics; DMD, discrete MD; TRA, turn A; TRB, turn B; CHC, central hydrophobic cluster; MHR, midhydrophobic region; CTR, C-terminal region.

†To whom correspondence should be addressed at: Physics Department, Center for Polymer Studies, Boston University, 590 Commonwealth Avenue, Boston, MA 02215. E-mail: brigita@bu.edu.

‡Present address: Department of Physics, Yeshiva University, 500 West 185th Street, New York, NY 10033.

mentation and the values of all of the model parameters are reported elsewhere (30).

The backbone hydrogen-bond interaction that normally occurs in proteins between the carbonyl oxygen of one amino acid and the amide hydrogen of another amino acid is implemented and is not amino acid-specific, as explained in detail by Ding *et al*. (30). The hydrogen-bond parameters are defined and their values given in Fig. 5 and Table 1, which are published as supporting information on the PNAS web site.

Amino Acid-Specific Interactions Caused by Side-Chain Hydropathy. The solvent is not explicitly present in our DMD approach. We introduce hydrophobic attraction/hydrophilic repulsion between pairs of side chains depending on the hydropathic nature of individual side chains. Hydrophobic attraction and hydrophilic repulsion both are implemented as effective interactions that mimic the effects of water or aqueous solution. In our model, the potential energy decreases when two hydrophobic residues interact, thus minimizing contacts with water. Conversely, the potential energy increases when two hydrophilic residues interact. This increase in energy thus favors noninteracting hydrophilic residues, which maximizes their contacts with water.

In our model we distinguish four types of side chains: hydrophobic, noncharged hydrophilic, charged hydrophilic, and neutral. There are different ways of implementing amino acid-specific hydropathic interactions. We chose the empirical amino acid hydropathy scale derived by Kyte and Doolittle (35). We consider amino acids Ile, Val, Leu, Phe, Cys, Met, and Ala to be hydrophobic, amino acids Asn, Gln, and His noncharged hydrophilic, and amino acids Arg, Lys, Asp, and Glu charged hydrophilic. The remaining amino acids with hydropathies below the threshold values are considered neutral. The amino acid-specific interactions are of two types: (*i*) an attractive interaction between two hydrophobic side chains; and (*ii*) a repulsive interaction between two noncharged hydrophilic side chains as well as between a charged hydrophilic and a noncharged hydrophilic side chain. The attractive and repulsive interactions caused by the hydropathic nature of side chains are implemented through a square-well potential between the pairs of side-chain beads $C_{\beta,i}$ and $C_{\beta,j}$ so that the two beads interact if the distance between their centers is less than the interaction range distance 0.75 nm. The potential energy of the effective attractive hydrophobic interaction is proportional to the mean of the relative hydrophobic strengths. The potential energy of the effective repulsive hydrophilic interaction is proportional to the mean of the relative hydrophilic strengths. Interaction strengths for specific pairs of amino acids are given explicitly in Table 2, which is published as supporting information on the PNAS web site. His, with a pK_a value of ≈6.0, is considered a noncharged hydrophilic amino acid because, at physiological conditions (pH = 7.4), only ≈4% of His is charged. Note that our model neglects the Coulombic interactions between two charged side chains because these interactions are effectively screened by the solvent and thus contribute less to controlling folding than do hydropathic interactions.

Results

The main hypothesis underlying our modeling approach is that the hydropathic nature of amino acids is the main driving force governing oligomerization. In the absence of hydropathic interactions, the only interactions allowed in the model are hard-core repulsion and hydrogen bonding. Such a model reproduces planar β-sheet oligomer conformations (32). By introducing the effective hydropathic interactions, the relative importance of the hydrogen-bond interaction is diminished, leading to less extended, more globular oligomer conformations. In the simulations described below, the maximum potential energy of the hydrophobic attraction (which occurs between two isoleucines) is set to 0.3 relative to the hydrogen-bond potential energy $E_{HB} = 1.0$. The temperature T is given in units of E_{HB}/k_B where k_B is the Boltzmann constant.

The primary structure of Aβ(1–42) is DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA. The amino acid sequence of Aβ40 is the first 40 amino acids of Aβ42. To present the results in a systematic manner, the peptide is segmented as follows: (*i*) Asp-1–Lys-16 is the N-terminal region; (*ii*) Leu-17–Ala-21 is the central hydrophobic cluster (CHC); (*iii*) Glu-22–Gly-29 is the turn A (TRA) region; (*iv*) Ala-30–Met-35 is the midhydrophobic region (MHR); (*v*) Val-36–Val-39 is the turn B (TRB) region; and (*vi*) Val-40 or Val-40–Ala-42 is the C-terminal region (CTR). Note that the CTR of Aβ40 consists of only one amino acid, Val-40.

Oligomer Size Distributions of Aβ40 and Aβ42. Aβ40 and Aβ42 oligomerize through different pathways (15). Determining the oligomer size-distribution differences between Aβ40 and Aβ42 *in silico* is a challenging and time-consuming task because it requires a relatively large set of trajectories to reach statistically significant conclusions. We present the results of eight different trajectories of Aβ40 and eight different trajectories of Aβ42. Each trajectory initially consists of 32 well separated monomer peptides enclosed in a cubic box of side 25 nm. This corresponds to a molar concentration of 3.4 mM, which is 10–100 times higher than the reported experimental concentrations of 30–300 μM (15, 17). We use this higher peptide concentration for the following reasons: (*i*) by increasing the concentration we increase the probability of intermolecular interactions, making the oligomerization process fast enough to be studied *in silico*; and (*ii*) with only 32 peptides per trajectory and a high peptide concentration, the number of monomers in our simulations is minimal, allowing us to obtain statistically reliable oligomer size distributions that include not only monomers and dimers but also higher-order oligomers.

We use the mostly α-helical Aβ monomer conformation from the Protein Data Bank (36) as a starting conformation (37, 38). We place 32 well separated peptides into our box and then perform a DMD simulation at a high temperature ($T = 2.0\ E_{HB}/k_B$), recording the 32-peptide conformation every 10,000 simulation steps. After 80,000 steps, we thus obtain eight different 32-peptide conformations, which we use as eight different starting conformations for eight trajectories. This initial process is done separately for Aβ40 and Aβ42. The initial conformations are characterized by a zero average potential energy and a secondary structure lacking α-helical or β-strand elements. Using the initial conformations described above, we then produce eight trajectories each for Aβ40 and Aβ42 at $T = 0.15\ E_{HB}/k_B$.

In Fig. 1 *Inset*, we show the time dependence of the average potential energy per peptide for Aβ40 and Aβ42. Each curve is an average over the eight corresponding trajectories. We present the potential energy of individual trajectories (black dots for Aβ40 and red dots for Aβ42) to show how much the potential energy per peptide varies from trajectory to trajectory. Initially, the potential energy of each trajectory is equal to zero, because all the peptides are in the initial zero-potential-energy conformation. As the monomers fold and oligomerization occurs, the potential energy per peptide decreases. After 4 million steps, each Aβ40 trajectory has a significantly higher potential energy per peptide compared to any Aβ42 trajectory. For clarity, we also plot the potential energy difference per peptide between Aβ42 and Aβ40 (green curve), which is equal to -4.2 ± 0.3 energy units.

We analyze the oligomer size distributions by analyzing the oligomer sizes of all eight trajectories of each alloform. Initially, Aβ40 and Aβ42 have the same size distributions with a peak at monomers. As the simulation progresses and peptides start to assemble into oligomers, the two distributions start to differ. The difference between Aβ40 and Aβ42 size distributions increases in significance with the simulation step. At $\approx 6 \times 10^6$ simulation steps, the difference between the two distributions reaches statistical significance ($P \leq 0.01$, χ^2 test), and this difference remains statistically significant for the rest of the simulation. At $\approx 8 \times 10^6$

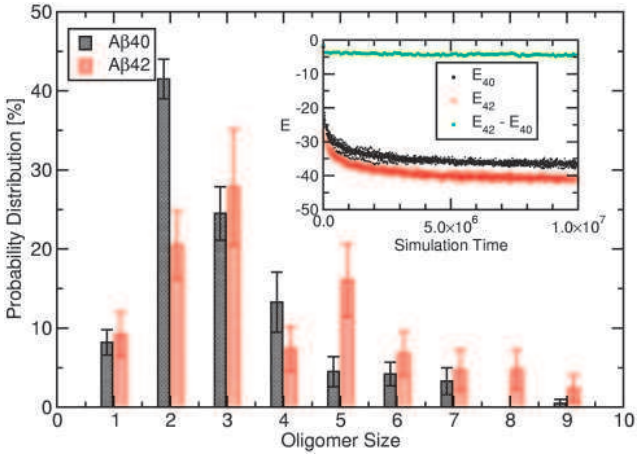

Fig. 1. Oligomer size distributions for Aβ40 and Aβ42, showing the occurrence probability [%] of monomers and oligomers. Standard errors of average probabilities are indicated by vertical bars. (*Inset*) The potential energy per peptide of Aβ40, E_{40} (black curve), and Aβ42, E_{42} (red curve), each averaged over eight trajectories. Potential energies of individual trajectories are shown as black (Aβ40) and red (Aβ42) dots. The difference between the average potential energies ($E_{42} - E_{40}$) is depicted by a green curve.

simulations steps, both size distributions reach a dynamic steady state, during which oligomers can break and assemble into oligomers of different sizes. Fig. 1 shows the oligomer size distributions of the two alloforms, each averaged over eight corresponding trajectories at a fixed simulation step and then averaged over three fixed simulation steps (9 million, 9.5 million, and 10 million steps). The probability of a particular oligomer size occurring in a given trajectory is determined as the number of oligomers of a given size divided by the total number of oligomers within the trajectory. The mean occurrence probabilities and their error bars in Fig. 1 are calculated by first finding the probabilities of individual trajectories and then calculating the means (the average occurrence probabilities) and their standard errors.

The mean occurrence probability for Aβ40 (Fig. 1, black histogram) peaks at dimers and monotonically decreases thereafter. The mean occurrence probability for Aβ42 (Fig. 1, red histogram) peaks near trimers and is followed by a significant decrease in tetramers. The mean occurrence probability for Aβ42 is characterized by another significant peak at pentamers and then monotonically decreases thereafter. This bimodal character of the Aβ42 probability distribution (as opposed to the monotonic decrease observed in the Aβ40 probability distribution) has been experimentally determined as the hallmark difference between the two alloforms (15). Comparing the Aβ40 and Aβ42 occurrence probabilities at a given oligomer size N, we find two significant differences: in dimers (significantly more abundant in Aβ40) and in pentamers (significantly more abundant in Aβ42). Time dependence of the number of monomers and oligomers in our simulations is presented in Fig. 6, which is published as supporting information on the PNAS web site.

Time Evolution of Contacts and Secondary Structure During Aβ Folding. To understand the mechanisms underlying the differences in the oligomer size distributions described above, we study the evolution of folding contacts and site-specific secondary structure in monomers, starting from unfolded initial conformations. By our definition, two beads are in contact whenever their centers of mass are closer than 0.75 nm. We take into account all of the beads of all amino acids in our contact map calculation. Contacts between pairs of amino acids can be visualized by using contact maps, 2-D representations of all amino acid pairs in contact in which contact frequencies are represented by different colors. The intramolecular contact map only takes into account the contacts between pairs of amino acids within the same peptide. The site-specific propensity for secondary structure formation is determined by using the STRIDE program (39, 40) within the VMD software package (41). The STRIDE program consists of the knowledge-based algorithm that uses hydrogen-bond energy and statistically derived backbone torsional angle information to return the secondary structure assignments in maximal agreement with crystallographers' designations.

We analyze 150–256 monomer conformations per fixed simulation step depending on how many monomers are present at that particular simulation step. The fixed simulation steps considered are at 0, 100, 1,000, and 10,000. Fig. 2 shows contact maps and secondary structure development during monomer folding of Aβ40 and Aβ42 from initially unfolded, high-temperature, and zero-potential-energy conformations. After 1,000 simulation steps, the contacts around Val-36–Val-39 develop, accompanied by a turn centered at Gly-33 (Aβ40) and a turn in the TRB region, which is prominent in Aβ42 but not in Aβ40. No significant β-strand structure is present. After 10,000 simulation steps, more contacts form further away from the contact Val-36–Val-39, accompanied by an additional turn in the TRA region. β-Strand structures also appear within the CHC and MHR (Aβ40) and within the CTR (Aβ42). After 100,000 simulation steps, contacts are formed between the N-terminal region on the one side and the CHC, MHR, and CTR on the other side. Turns centered at Gly-9, Gly-25, Gly-29, and Gly-33 appear in both Aβ40 and Aβ42 with similar propensities. The turn in the TRB region remains prominent in Aβ42 but not in Aβ40. Aβ40 and Aβ42 monomers share the β-strand structure within the CHC and the MHR. Aβ40 has a prominent β-strand structure at Ala-2–Phe-4 that is not present in Aβ42. Aβ42 has significantly more β-strand structure at Glu-11–His-14 than does Aβ40. In addition, Aβ42 has a strong β-strand structure at Val-39–Val-40 that is not present in Aβ40.

Secondary, Tertiary, and Quaternary Structure of Aβ Oligomers. Here we study how the secondary, tertiary, and quaternary structure of Aβ monomers and oligomers depend on the assembly state after 9 million simulation steps when the monomers and oligomers are in a quasi-steady state. The secondary structure of Aβ oligomers is obtained once again by using the STRIDE program as described above. Tertiary and quaternary structures are obtained by analyzing the intramolecular and intermolecular contact maps. The intramolecular contact map contains information about the tertiary structure of peptides within the oligomers. The intermolecular contact map only takes into account the contacts between pairs of amino acids that belong to different peptides and thus yields information on the way different peptides assemble into the oligomer under consideration (the quaternary structure). We first decompose each trajectory at 9 million, 9.5 million, and 10 million steps into individual monomer, dimer, trimer, tetramer, and pentamer conformations and then analyze each oligomer assembly state separately. Results are presented in detail in *Supporting Text*, which is published on the PNAS web site.

Using the STRIDE program, we obtain the site-specific propensities for a turn, a β-strand, and an α-helix at a given oligomer size N. These site-specific propensities are averaged over N peptides within each individual N-mer conformation and then averaged over all N-mer conformations. Our results show that the α-helix propensity is zero along the whole peptide for all oligomer sizes in both alloforms. We find significant differences between the two alloforms in the turn and β-strand propensities. Details of our analysis with the graph showing site-specific turn and β-strand propensities per oligomer size are given in Fig. 7, which is published as supporting information on the PNAS web site. Comparing the turn propensities of the two alloforms, we find the main difference between the two alloforms is in the TRB region: in Aβ42, a strong

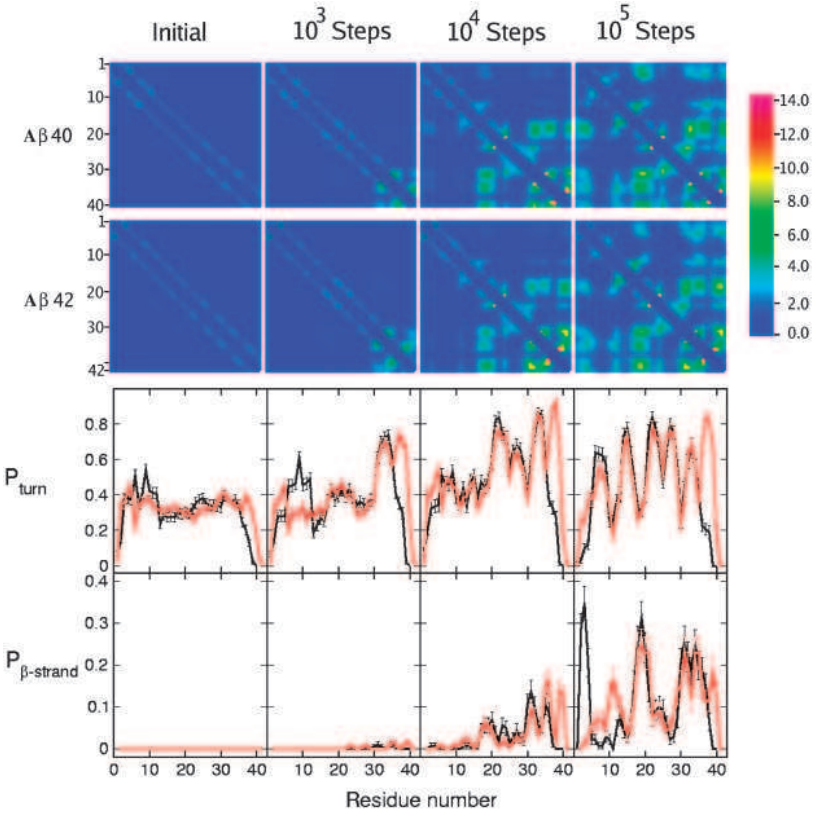

Fig. 2. Time evolution of contacts and secondary structure elements during monomer folding from high-temperature, zero-potential-energy initial conformations. Columns correspond to states at the start of the simulation (initial conformations) and after 10^3, 10^4, and 10^5 simulation steps. (*Upper*) The contact maps for Aβ40 and Aβ42 are averages of >150 monomer conformations each. (Asp-1, Asp-1) is at the upper-left corner of the contact maps and (Val-40, Val-40) for Aβ40 or (Ala-42, Ala-42) for Aβ42 is at the lower-right corner. The strength of the contact is color-coded following the rainbow scheme: from blue (no contact), through green, yellow, and orange, to red (strongest contact), as shown on the bar with the scale on the right. (*Lower*) Time evolution of the turn propensity P_{turn} and the β-strand propensity $P_{\beta\text{-strand}}$ is presented. The black curves correspond to Aβ40, and the red ones correspond to Aβ42. Error bars indicate the SEM values.

turn is equally present in monomers and oligomers, whereas in Aβ40, a turn appears in dimers and then gets more pronounced as the oligomer size increases (see Fig. 7). This result suggests that the turn in the TRB region plays an important role in Aβ assembly. Comparing the β-strand propensities of the two alloforms, we find that the peptide region Ala-2–Phe-4 in Aβ40 has a pronounced β-strand structure, absent in Aβ42. This difference is present in all assembly states. In addition, a significant β-strand structure exists at Val-39–Val-40 in all Aβ42 oligomers (dimers through pentamers) but not in Aβ40 oligomers.

We then investigate the assembly-state dependence of tertiary and quaternary structure elements by analyzing the intramolecular and intermolecular contact maps. Details of this analysis with a figure showing intramolecular and intermolecular contact maps per assembly state are given in Fig. 8, which is published as supporting information on the PNAS web site. Intramolecular contact map analysis shows that the turn in the TRB region brings Met-35 (as well as Ala-30, Ile-31, Ile-32, and Leu-34) in contact with the CTR. Although these contacts are strong in Aβ42 pentamers, they are barely present in Aβ40 pentamers, presumably because of the absence of the strongly hydrophobic amino acid Ile-41 in the latter. These results imply that the role of Met-35 (and neighboring side chains) might be quite different in Aβ40 and Aβ42, which is in agreement with recent *in vitro* findings elucidating the role of Met-35 in paranucleus formation (17). Intermolecular contact map analysis shows that in Aβ40, the strongest contacts are connecting pairs of CHC regions, whereas the contacts between the C-terminal Ile-41 and the CHC dominate in Aβ42. Moreover, the relative importance of the CTR for intermolecular contact formation in Aβ42 increases with the oligomer size and is strongest in a pentamer. This observation is in agreement with the experimental results, indicating the importance of Ile-41 for paranucleus formation in Aβ42 (16).

Experimental results (15, 16) indicate that it is the formation of paranuclei (i.e., pentamers and hexamers of Aβ42) that underlies the differences between Aβ40 and Aβ42 oligomerization. In Fig. 3, the intramolecular contact maps of pentamer conformations are presented. Both Aβ40 and Aβ42 display a turn centered at Gly-25–Ser-26 (black squares in Fig. 3). In addition, the Aβ42 contact map contains a significantly greater number of contacts centered around the strongest contact, Val-36–Val-39, than does the contact map of Aβ40 (red squares in Fig. 3). The intramolecular contact maps of Aβ40 and Aβ42 pentamers in the region Leu-17–Val-40/Ala-42 are presented in a more explicit form in Fig. 3 *c* and *d*. Substantial differences are apparent, particularly involving Met-35. In Aβ40, Met-35 is in contact with the CHC (Leu-17–Ala-21) but not with the C terminus (in particular Val-39 and Val-40), whereas in Aβ42, there are significantly more contacts between Met-35 and the C terminus (Val-39, Val-40, Ile-41, and Ala-42) in addition to the contacts between Met-35 and the CHC (Leu-17–Ala-21). Analogous differences between Aβ40 and Aβ42 occur for other amino acids in the proximity of Met-35, including Ile-31, Ile-32, and Leu-34.

Geometrical Characteristics of Aβ Pentamers. Typical pentamers of Aβ40 and Aβ42 are presented in Fig. 4 *a* and *b*. Pentamers of both Aβ40 and Aβ42 are globular and have their C termini within the assembly core and their N termini on the surface. A significant difference between the Aβ40 and Aβ42 assemblies is that the N termini of the Aβ42 pentamers are more extended and less structured. To quantify this difference, we calculate the distribution of distances between C_α atoms of Asp-1 and Val-40 in both Aβ40 and Aβ42 pentamers. The two distributions differ significantly: whereas in Aβ40 the most probable distance is $\approx 1.5 \pm 0.5$ nm, the distribution in Aβ42 pentamers does not have a well defined peak and spans larger distances between 1.5 and 3.5 nm. We obtain a similar result if we use C_β atoms instead of C_α atoms or if the distance Asp-1–Ala-42 is used in the analysis (data not shown).

Next we investigate whether this difference in N-terminal organization affects the mass distribution within the pentamer. To do so,

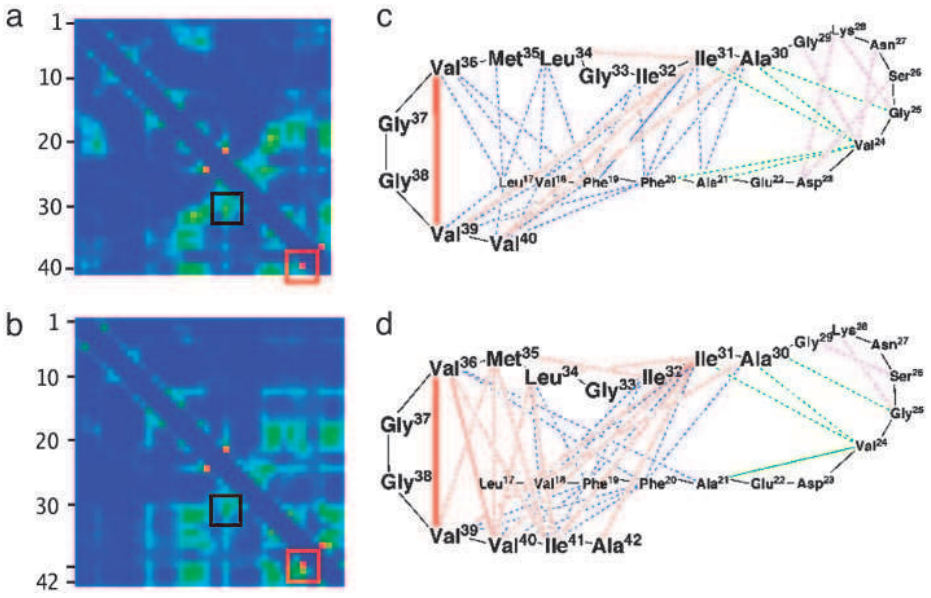

Fig. 3. Intramolecular contacts in pentamers. (*a* and *b*) Average contact maps of intramolecular contacts within a pentamer conformation. The averages are calculated by using individual contact maps of 11 Aβ40 and 34 Aβ42 pentamers using three fixed simulation steps (9 million, 9.5 million, and 10 million simulation steps). *a* and *c* correspond to Aβ40, and *b* and *d* correspond to Aβ42. (Asp-1, Asp-1) is at the upper-left corner of the contact maps and (Val-40, Val-40) for Aβ40 or (Ala-42, Ala-42) for Aβ42 is at the lower-right corner. The strength of the contacts is color-coded as described for Fig. 2. The centers of the TRA and TRB regions are marked by black and red squares, respectively. (*c* and *d*) Diagrams of the most important intramolecular contacts within a pentamer calculated by using the contact map data from *a* and *b*. Solid lines represent stronger contacts, and dashed lines represent weaker contacts. Color code: magenta, contacts inside the MHR; green, contacts between the MHR and the CHC or the CTR; blue, contacts between the CHC and the CTR; red, contacts inside the CTR.

we calculate the average number of atoms per unit volume as a function of the radial distance from the center of mass. In this calculation, we take into account all the atoms within pentamers except hydrogens. The result is shown in Fig. 4*d Inset*. We conclude that both Aβ40 and Aβ42 have a relatively constant atom number density up to 1 nm, and at larger radial distances, the atom number density decays monotonically. There is no significant difference between the overall size and mass distribution within Aβ40 and Aβ42 pentamers. As the radial density goes to zero between 2 and 3 nm, we can estimate that the pentamers are of an average diameter 5 ± 1 nm. This result agrees well with the diameter of micelle-like intermediates and/or stable globular oligomers as experimentally determined by small-angle neutron scattering (42) and atomic force microscopy (43).

Distances of the C_α atoms of individual amino acids from the center of mass of a pentamer are also calculated, then averaged over all the peptides in the pentamer, and finally over all pentamer conformations (Fig. 4*d*). Aβ40 and Aβ42 both display three peptide regions particularly close to the center of mass: CHC, MHR, and CTR. These three regions all are strongly hydrophobic, and thus it is reasonable that they would form the core of the pentamer. There is a significant difference between the two alloforms within the region Asp-1 through His-6. In Aβ40 pentamers, the average distance of these residues from the center of mass is 1.8–2.0 nm, whereas in Aβ42 pentamers, these distances are considerably larger (2.2–2.5 nm). In Aβ40 and Aβ42, the peptide region Asp-1–His-6 is the farthest away from the center of mass, consistent with its hydrophilic nature. This effect is more pronounced in Aβ42. In both alloforms, other parts of the peptide relatively far from the center of mass are His-13–His-14, Ser-26–Lys-28, and the TRA region (radial distance >1.5 nm) as well as Glu-22–Asp-23 and Gly-37–Gly-38 (radial distance >1.3–1.4 nm). These results are consistent with experimental studies and first principles of protein folding and assembly.

Discussion and Conclusions

Aβ oligomerization seems to be a seminal event in the pathogenesis of AD. Understanding the Aβ oligomerization process thus has been an important goal for those seeking to develop therapeutic agents to combat the disease. Unfortunately, Aβ self-association is a complicated process involving a number of folding and assembly pathways. Stable intermediates do not form, making identification and study of key pathologic aggregates extremely difficult. In addition, early conformers that form on pathways producing toxic assemblies exist in amounts and with lifetimes far too short for conventional experimental observation. For these reasons, here we

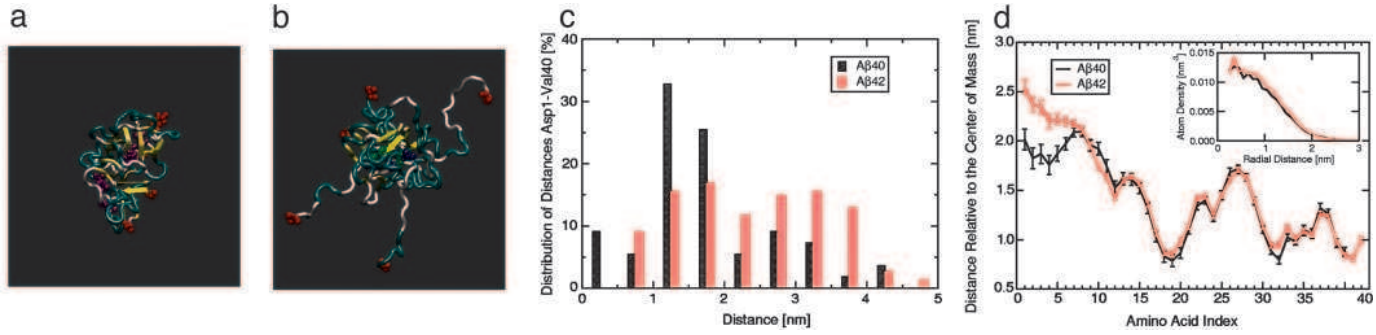

Fig. 4. Structural features of pentamers. Typical Aβ40 (*a*) and Aβ42 (*b*) pentamers. The secondary structure of pentamers is shown as a silver tube (random coil-like structure), light-blue tube (turn), and yellow ribbon (β-strand). Red spheres in both *a* and *b* represent the N-terminal Asp-1. (*a*) The C-terminal amino acids Val-39 and Val-40 are shown in purple. (*b*) The C-terminal amino acid Ile-41 is shown in green, and Ala-42 is shown in blue. (*c*) Distribution of intramolecular distances between Asp-1 and Val-40 within Aβ40 (black) and Aβ42 (red) pentamers. The distributions are significantly different ($P < 10^{-4}$, χ^2 test). (*d*) The average distances from the center of mass of Aβ40 and Aβ42 pentamers per residue. The error bars represent SEM values. (*d Inset*) The atom density (number of atoms per volume unit) in dependence on the radial distance from the center of mass for pentamers of Aβ40 (black) and Aβ42 (red).

apply *in silico* methods to simulate the folding and oligomerization of Aβ. We use a coarse-grained Aβ model with no explicit solvent and an efficient DMD algorithm. This approach allows the study of systems of ≥32 peptides of molecular mass ≈ 4.5 kDa. Analysis of the data produced in simulations of Aβ40 and Aβ42 folding and oligomerization reveals structural and kinetic features relevant to understanding the distinct biophysical and biological behaviors of the two peptides *in vivo*. Our results show that the oligomer size distributions of Aβ40 and Aβ42 differ significantly. Aβ40 produces more dimers than Aβ42, and conversely, more pentamers are formed by Aβ42 than by Aβ40. These observations are consistent with *in vitro* oligomer size-distribution studies that reveal a monotonic decrease of the Aβ40 size distribution and a bimodal Aβ42 distribution peaked at pentamers/hexamers (15).

Our study of Aβ40 and Aβ42 monomer folding and oligomerization indicates that structural differences between the two peptides occur as early as the monomer folding stage. In particular, a folded Aβ42 monomer contains a turn centered at Gly-37–Gly-38 that is almost absent in a folded Aβ40 monomer. This observation agrees with preliminary *in vitro* results on monomer folding obtained by using limited proteolysis, LC-MS, and NMR (N. D. Lazo, M. A. Grant, M. C. Condron, A. C. Rigby, and D.B.T., unpublished data). Furthermore, our *in silico* data show that the initial folding event in Aβ monomer folding involves contacts between Val-36–Val-39 and their neighboring amino acids. Folding in the Ala-21–Ala-30 seems to occur later. Experimental work has not yet addressed the temporal order of these folding events.

Our structural analysis shows that the core of the Aβ pentamer is made up primarily of amino acids Leu-17–Ala-21, Ala-30–Met-35, and Val–40/Val-40–Ala-42, where hydrophobic amino acids are concentrated, in agreement with the basic principles of globular protein organization. The N termini of both alloforms are the peptide regions most likely found at the surface of pentamers but significantly more so in Aβ42 pentamers. The N termini of Aβ40 peptides within pentamers are more spatially restricted than are the N termini of Aβ42 peptides within pentamers. Aβ40 displays a β-strand structure involving Ala-2–Phe-4 that is not present in Aβ42. This structural element is present in all Aβ40 assemblies. Because the hydrophilic N termini of Aβ40 and Aβ42 are on the surface of oligomers, the presence of the N-terminal β-strand in Aβ40 may shield the hydrophobic core of the oligomer. Because removal of this "shielding" is necessary for intermolecular interactions among hydrophobic cores of multiple oligomers, higher-order association reactions may be energetically unfavorable. This conclusion suggests that it is this structural difference between Aβ40 and Aβ42 that is responsible for the distinct oligomer size distributions of the two peptides. If so, then a substitution of the hydrophobic N-terminal residues Ala-2 and/or Phe-4 in Aβ40 by a polar, uncharged amino acid should shift the Aβ40 oligomerization characteristics toward those of Aβ42. This hypothesis is amenable to experimental verification.

In our model, the origin of the oligomer size-distribution difference between Aβ40 and Aβ42 is the additional hydrophobicity provided by the two C-terminal amino acids of Aβ42: Ile-41 and Ala-42. Ile-41, the most hydrophobic amino acid in the Kyte and Doolittle tabulation (35), plays an especially important role in this regard. This result is in agreement with *in vitro* studies that show that the addition of Ile-41 to Aβ40 is sufficient to induce the formation of paranuclei but insufficient to support paranucleus self-association (15). The results of our *in silico* study: (*i*) are consistent with experimental data, suggesting that the approach is biologically and clinically relevant; (*ii*) reveal features of the assembly process unobservable by other methods; (*iii*) provide experimentally testable hypotheses about Aβ folding and assembly; and (*iv*) provide a method for *in silico* testing of therapeutic compounds through inclusion of these compounds with Aβ monomers during simulations and observation of the effects on Aβ folding and oligomerization.

We thank F. Ding for implementation of the four-bead protein model and many helpful discussions. This work was supported by the Memory Ride Foundation and National Institutes of Health Grants NS38328, NS44147, and AG18921 (to D.B.T.), the Foundation for Neurologic Diseases (D.B.T.), and American Federation for Aging Research Grant A04084 (G.B.).

1. Walsh, D. M., Lomakin, A., Benedek, G. B., Condron, M. M. & Teplow, B. D. (1997) *J. Biol. Chem.* **272,** 22364–22372.
2. Hartley, D. M., Walsh, D. M., Ye, C. P., Diehl, T., Vasquez, S., Vassilev, P. M., Teplow, D. B. & Selkoe, D. J. (1999) *J. Neurosci.* **19,** 8876–8884.
3. Harper, J. D., Lieber, C. M. & Lansbury, P. T., Jr. (1997) *Chem. Biol.* **4,** 951–959.
4. Lambert, M.P., Barlow, A.K., Chromy, B. A., Edwards, C., Freed, R., Liosatos, M., Morgan, T. E., Rozovsky, I., Trommer, B., Viola, K. L., *et al*. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6448–6453.
5. Hsia, A. Y., Masliah, E., McConlogue, L., Yu, G. Q., Tatsuno, G., Hu, K., Kholodenko, D., Malenka, R. C., Nicoll, R. A. & Mucke, L. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3228–3233.
6. El-Agnaf, O. M., Mahil, D. S., Patel, B. P. & Austen, B. M. (2000) *Biochem. Biophys. Res. Commun.* **273,** 1003–1007.
7. Mucke, L., Masliah, E., Yu, G.-Q., Mallory, M., Rockenstein, E. M., Tatsuno, G., Hu, K., Kholodenko, D., Johnson-Wood, K. & McConlogue, L. (2000) *J. Neurosci.* **20,** 4050–4058.
8. El-Agnaf, O. M., Nagala, S., Patel, B. P. & Austen, B. M. (2001) *J. Mol. Biol.* **310,** 157–168.
9. Dodart, J.-C., Bales, K. R., Gannon, K. S., Greene, S. J., DeMattos, R. B., Mathis, C., DeLong, C. A., Wu, S., Wu, X., Holtzman, D. M. & Paul, S. M. (2002) *Nat. Neurosci.* **5,** 452–457.
10. Kirkitadze, M. D., Bitan, G. & Teplow, D. B. (2002) *J. Neurosci. Res.* **69,** 567–577.
11. Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J. & Selkoe, D. J. (2002) *Nature* **416,** 535–539.
12. Westerman, M. A., Cooper-Blacketer, D., Mariash, A., Kotilinek, L., Kawarabayashi, T., Younkin, L. H., Carlson, G. A., Younkin, S. G. & Ashe, K. H. (2002) *J. Neurosci.* **22,** 1858–1867.
13. Gong Y., Chang, L., Viola, K. L., Lacor, P. N., Lambert, M. P., Finch, C. E., Krafft, G. A. & Klein, W. L. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 10417–10422.
14. Bitan, G., Lomakin, A. & Teplow, D. B. (2001) *J. Biol. Chem.* **276,** 35176–35184.
15. Bitan, G., Kirkitadze, M. D., Lomakin, A., Vollers, S. S., Benedek G. B. & Teplow, D. B. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 330–335.
16. Bitan, G., Vollers, S. S. & Teplow, D. B. (2003) *J. Biol. Chem.* **278,** 34882–34889.
17. Bitan, G., B. Tarus, Vollers, S. S., Lashuel, H. A., Condron, M. M., Straub, J. E. & Teplow, D. B. (2003) *J. Am. Chem. Soc.* **125,** 15359–15365.
18. Zagrovic, B., Snow, C. D., Shirts, M. R. & Pande, V. S. (2002) *J. Mol. Biol.* **323,** 927–937.
19. Snow, C. D., Nguyen, N., Pande, V. S. & Gruebele, M. (2002) *Nature* **420,** 102–106.
20. Daidone, I., Simona, F., Roccatano D., Broglia, R. A., Tiana, G., Colombo, G. & Di Nola, A. (2004) *Proteins* **57,** 198–204.
21. Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W. & Glabe, C. G. (2003) *Science* **300,** 486–489.
22. Zhou, Y. & Karplus, M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 14429–14432.
23. Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (1998) *Fold. Des.* **3,** 577–587.
24. Zhou, Y. & Karplus, M. (1999) *J. Mol. Biol.* **293,** 917–951.
25. Smith, A. V. & Hall, C. K. (2001) *Proteins Struct. Funct. Genet.* **44,** 344–360.
26. Smith, A. V. & Hall, C. K. (2001) *J. Mol. Biol.* **312,** 187–202.
27. Borreguero, J. M., Dokholyan, N. V., Buldyrev, S. V., Shakhnovich, E. I. & Stanley, H. E. (2002) *J. Mol. Biol.* **318,** 863–876.
28. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002) *Biophys. J.* **83,** 3525–3532.
29. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002) *J. Mol. Biol.* **324,** 851–857.
30. Ding, F., Borreguero, J. M., Buldyrev, S. V., Stanley, H. E. & Dokholyan, N. V. (2003) *Proteins Struct. Funct. Genet.* **53,** 220–228.
31. Peng, S., Ding, F., Urbanc, B., Buldyrev, S. V., Cruz, L., Stanley, H. E. & Dokholyan, N. V. (2004) *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **69,** 041908.
32. Urbanc, B., Cruz, L., Ding, F., Sammond, D., Khare, S., Buldyrev, S. V., Stanley, H. E. & Dokholyan, N. V. (2004) *Biophy. J.* **87,** 2310–2321.
33. Rapaport, D. C. (1997) *The Art of Molecular Dynamics Simulation* (Cambridge Univ. Press, Cambridge, U.K.).
34. Takada, S., Luthey-Schulten, Z. & Wolynes, P. G. (1999) *J. Chem. Phys.* **110,** 11616–11629.
35. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
36. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235–242.
37. Coles, M., Bicknell, W., Watson, A. A., Fairlie, D. P. & Craik, D. J. (1998) *Biochemistry* **37,** 11064–11077.
38. Crescenzi, O., Tomaselli, S., Guerrini, R., Salvatori, S., D'Ursi, A. M., Temussi, P. A. & Picone, D. (2002) *Eur. J. Biochem.* **269,** 5642–5648.
39. Frishman, D. & Argos, P. (1995) *Proteins Struct. Funct. Genet.* **23,** 566–579.
40. Heinig, M. & Frishman, D. (2004) *Nucleic Acids Res.* **32,** W500–W502.
41. Humphrey, W., Dalke, A. & Schulten, K. (1996) *J. Mol. Graphics* **14,** 33–38.
42. Yong, W., Lomakin, A., Kirkitadze, M. D., Teplow, D. B., Chen, S. H. & Benedek, G. B. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 150–154.
43. Chromy, B. A., Nowak, R. J., Lambert, M. P., Viola, K. L., Chang, L., Velasco, P. T., Jones, B. W., Fernandez, S. J., Lacor, P. N., Horowitz, P., *et al*. (2003) *Biochemistry* **42,** 12749–12760.