Simulation study on the disordered state of an Alzheimer's β amyloid peptide A β (12–36) in water consisting of random-structural, β -structural, and helical clusters

JINZEN IKEBE,¹ NARUTOSHI KAMIYA,² JUN-ICHI ITO,¹ HEISABURO SHINDO,³
AND JUNICHI HIGO¹

¹School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, Hachioji, Tokyo 192-0392, Japan

²Clinical Genome Informatics Center, Kobe University, Graduate School of Medicine, Chuo-ku, Kobe 650-0047, Japan

³School of Pharmacy, Tokyo University of Pharmacy and Life Sciences, Hachioji, Tokyo 192-0392, Japan

(RECEIVED December 12, 2006; FINAL REVISION April 30, 2007; ACCEPTED May 14, 2007)

Abstract

The monomeric Alzheimer's β amyloid peptide, A β , is known to adopt a disordered state in water at room temperature, and a circular dichroism (CD) spectroscopy experiment has provided the secondary-structure contents for the disordered state: 70% random, 25% β -structural, and 5% helical. We performed an enhanced conformational sampling (multicanonical molecular dynamics simulation) of a 25-residue segment (residues 12–36) of A β in explicit water and obtained the conformational ensemble over a wide temperature range. The secondary-structure contents calculated from the conformational ensemble at 300°K reproduced the experimental secondary-structure contents. The constructed free-energy landscape at 300°K was not plain but rugged with five clearly distinguishable clusters, and each cluster had its own characteristic tertiary structure: a helix-structural cluster, two β -structural clusters, and two random-structural clusters. This indicates that the contribution from the five individual clusters determines the secondary-structure contents experimentally measured. The helical cluster had a similarity with a stable helical structure for monomeric A β in 2,2,2-trifluoroethanol (TFE)/water determined by an NMR experiment: The positions of helices in the helical cluster were the same as those in the NMR structure, and the residue–residue contact patterns were also similar with those of the NMR structure. The cluster–cluster separation in the conformational space indicates that free-energy barriers separate the clusters at 300°K. The two β -structural clusters were characterized by different strand–strand hydrogen-bond (H-bond) patterns, suggesting that the free-energy barrier between the two clusters is due to the H-bond rearrangements.

Keywords: free-energy landscape; generalized ensemble; multicanonical molecular dynamics; all-atom model; explicit water; fibril formation

Reprint requests to: Junichi Higo, School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan; e-mail: higo@ls.toyaku.ac.jp; fax: +81-426-76-5863.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.062721907>.

Protein tertiary structure varies depending on the solvent condition. Many proteins have their activities under specific solvent conditions, and the activities change or are lost in different solvent. This fact indicates that alternation of the solvent condition varies the free-energy

landscape with shifting of the thermodynamically stable tertiary structure from one region to another in the conformational space. In comparison with proteins, short peptides are more affected by solvent. Generally, a peptide in water forms hydrogen bonds (H-bonds) with water molecules and tends to adopt random structures. Contrarily, the peptide in organic solvent is stabilized by intra-polypeptide H-bonds and tends to involve secondary-structure elements, such as α -helix and β -strand. Therefore, the conformational ensembles of the peptide in the different solvents are generally different (Wei and Shea 2006). However, would these ensembles be completely unrelated to each other? Particularly, in the free-energy landscape of the disordered state in water, is there any structural seed that can mature to an ordered state in a different solvent?

To answer the above questions, an Alzheimer's β amyloid peptide (A β) is an appropriate system. The A β peptide originally consists of 40 (or 42) amino acid residues and is known as a main cause of Alzheimer's disease. Generally, a fragment consisting of residues m – n cut from the full-length A β is expressed as A $\beta(m$ – n). It is known that the monomeric A β adopts various tertiary structures depending on the solution condition. It has been experimentally proved that the monomeric A β is disordered in water at room temperature from a CD experiment (Fezoui and Teplow 2002; Chen et al. 2006): The fraction for random conformations is $\sim 70\%$, although small fractions for β - and α -structures are involved in the disordered phase ($\sim 25\%$ for the β -structures and $\sim 5\%$ for the α -structures). This experimental measurement may suggest that the helical and strand regions may be thermally generated/annihilated in the monomeric A β in water. However, it is not known from the experiment whether the random, α -, and β -structures are distributed plainly in the conformational space or distinctly as different conformational clusters.

Contrarily, the monomeric A β in a 2,2,2-trifluoroethanol (TFE)/water mixture solution has a higher helix content than A β in water has (Fezoui and Teplow 2002; Chen et al. 2006). Particularly, A β adopts a unique and stable helix-rich structure (PDB ID, 1AML) in 40% (v/v) TFE/water as determined by an NMR experiment (Sticht et al. 1995). This NMR structure consists of helix 1 (residues 15–23) and helix 2 (residues 31–35) with a coiled region (residues 24–30) connecting the two helices, while the terminal regions (residues 1–11 and 37–40) are disordered. Another unique structure of A β is β -structures observed in the aggregation (i.e., amyloid fibril) (Petkova et al. 2002; Sciarretta et al. 2005). However, unlike the β -structures mentioned above, no stable β -structure has been reported at an atomic resolution for monomeric A β . Thus, we do not know the origin of the 25% β content involved in the disordered A β in water. Here, we address three questions to understand the

disordered A β : (1) How are the secondary-structure contents realized in the free-energy landscape? (2) Does the conformational ensemble involve a structure that has a similarity to the structures in 40% TFE/water? (3) Does the 25% β content exist as an isolated cluster in the free-energy landscape? Experiments provided the secondary-structure contents for the monomeric A β in water, as explained above. However, structural details of this peptide concerning these questions can be obtained only from simulations.

Powerful conformational sampling methods, such as replica-exchange molecular dynamics (REMD) simulation (Sugita and Okamoto 1999) or multicanonical MD (McMD) simulation (Nakajima et al. 1997), essentially assist the investigation of an accurate free-energy landscape. The McMD simulations (Higo et al. 2001; Kamiya et al. 2002; Ikeda and Higo 2003; Ikeda et al. 2003) have sampled throughout the conformations of short peptides about 10 residues long in explicit water starting from completely random conformations (i.e., no information from the native structures of the peptides was used in the simulations) and obtained their free-energy landscapes. An REMD simulation was used to study monomeric A β in explicit water and showed that A β was disordered in water (Baumketner and Shea 2007). However, no simulation work has yet reproduced the experimentally measured secondary-structure contents. Therefore, the three questions proposed above have not yet been solved. We consider solving these questions to be essentially important for understanding the disordered state of A β .

To answer the questions, we performed an enhanced conformational sampling, McMD simulation, of A β (12–36) in explicit water. The McMD method enhances the conformational sampling, so that the conformation can randomly fluctuate in an energy space without being trapped into energy minima, and provides a thermodynamic ensemble (i.e., the canonical ensemble) of conformations at any temperature. We constructed an accurate free-energy landscape from the canonical ensemble at 300°K from the McMD simulation trajectory. We report that the computed secondary-structure contents coincide with the CD-experimental ones. By analyzing the free-energy landscape, we solve the three questions.

Results

Selection of force field and secondary-structure contents of A β (12–36) in water

One of the main purposes for the current study is to investigate the relation between the computationally sampled structures of A β in water and the 1AML structure (Sticht et al. 1995) in 40% TFE/water, as explained in the introduction. Thus, the unstructured terminal regions (residues 1–11 and 37–40) of A β (1–40) in the

1AML structure are beyond the scope of the current study, so we removed the terminal regions. To determine an appropriate force field, we carried out eight McMD simulations using eight different AMBER-based hybrid force fields $E_{\text{hybrid}}(w)$, varying the weight w for mixing the AMBER parm94 and parm96 force-field parameters. We then calculated the secondary-structure contents in the conformational ensemble equilibrated at 300°K for each of the eight McMD simulations. Details for preparation of the system, simulation technique, and data analyses are given in Materials and Methods.

Figure 1A shows the secondary-structure contents of A β (12–36) at 300°K for the eight force fields. We recognize at a glance that $E_{\text{hybrid}}(0.70)$ is the best force field of the eight to reproduce the measured values from

Figure 1. Secondary-structure contents vs. weight parameter, w , in the force field $E_{\text{hybrid}}(w)$: (A) A β (12–36); (B) full-length A β (1–40). Conformational ensemble equilibrated at 300°K was used to compute the secondary-structure content. Thick solid, broken, and dotted lines indicate secondary-structure contents for helix, β -, and random structure contents, respectively. In both A and B, horizontal lines indicate helix (solid line), β -structure (broken line), and random structure (dotted line) contents as measured from CD experiments of A β (1–40) in water (Fezoui and Teplow 2002).

the CD experiment (Fezoui and Teplow 2002). The experimental secondary-structure contents were given in the introduction. Those from $E_{\text{hybrid}}(0.70)$ were 67.8% for the random structures, 26.7% for the β -structures (26.7%), and 5.5% for the α -structures.

The segment A β (12–36) for our simulation is different from the full-length A β (1–40) for the CD experiment. As is well known, TFE promotes a polypeptide to adopt some secondary structure. Even in 40% TFE/water, the terminal regions (residues 1–11 and residues 37–40, which are the removed regions in our simulation) of the monomeric A β (1–40) are highly disordered, as shown in 1AML structure. Thus, it is likely that the terminal regions are disordered in water. We then corrected the secondary-structure contents from the values for A β (12–36) to those for the full-length A β (1–40), assuming that the secondary-structure contents for residues 12–36 are the same as those from the current simulation and that the terminal regions are unstructured. Figure 1B indicates that $E_{\text{hybrid}}(0.70)$ is again the best of the eight force fields: The secondary-structure contents are 79.9%, 16.7%, and 3.4% for the random, β -, and helical structures, respectively. In Figure 1, A and B, the helix content drastically decreased when w was varied from 0.65 to 0.70.

Based on the above analysis, we focus our attention on the conformational ensemble from $E_{\text{hybrid}}(0.70)$. All the data analyzed below are those from $E_{\text{hybrid}}(0.70)$, although we do not necessarily mention the force field used.

Clusters in the conformational space

The McMD trajectory generated a 300°K conformational ensemble consisting of 841 snapshots using the reweighting technique (see Materials and Methods). To investigate the distribution of the sampled structures in a conformational space, we applied principal component analysis (PCA, see Materials and Methods) to the ensemble at 300°K. Figure 2 indicates the sampled structures projected on the PCA subspace constructed by the first to third eigenvectors, where we found five clusters and named them Cluster A–E. The cluster–cluster separation in the figure indicates the existence of free-energy barriers at 300°K.

To characterize each cluster, the structures were colored depending on the secondary-structure contents (see Fig. 2 legend). Remember that the majority of the sampled structures were assigned to random structures (random-structure content = 67.8%), and all of the clusters involve random structures (white dots in Fig. 2). However, each cluster clearly has its own characteristics. Clusters A and B involve β -rich structures more than the others do, although they also contain helical structures. Cluster C involves helix-rich structures more than the others do, although it also contains β -rich structures. Clusters D and E are random-structural clusters. The structural variety shown in Figure 2 indicates

Figure 2. Distribution of 841 structures of A β (12–36) sampled at 300°K, which are projected on the PCA subspace constructed by the first (PC1), second (PC2), and third (PC3) eigenvectors. Clusters A–E are shown by red characters. Structures are colored depending on their secondary-structure contents. The most helical structures involved 11 helical residues except for Ace and Nme: $N_{\max}^{\text{helix}} = 11$. Red and magenta dots represent structures with helical residues of 65% or more of $N_{\max}^{\text{helix}}$ and those of 50%–65% of $N_{\max}^{\text{helix}}$, respectively. The most β -rich structures involved 20 β -structural residues: $N_{\max}^{\text{strand}} = 20$. Blue and light blue dots represent structure with β -structural residues of 80% or more of $N_{\max}^{\text{strand}}$ and those of 65%–80% of $N_{\max}^{\text{strand}}$, respectively. The other structures were colored by white. Contours construct isodensity surfaces introduced to clearly show the five clusters.

that A β in water is disordered as a whole without adopting only a single structure. A small number of structures are not assigned to any cluster in Figure 2. These structures have intermediate characteristics between the clusters.

Helical Cluster C

Cluster C mainly consists of random and helix-rich structures. Figure 3A shows the helix probability at each residue site in Cluster C, the other clusters, and the 1AML structure. The figure shows that the helical regions in the 1AML structure correlate with those in Cluster C. Figure 3, B and C, displays a helix-rich structure in Cluster C and an NMR model of the 1AML structure, respectively. In contrast, the helix probability from the other clusters (see broken line in Fig. 3A) is considerably different from the helix regions in the 1AML structure. Thus, only Cluster C has similarity to the 1AML structure.

On the other hand, one may notice in Figure 3A that residues 15,16, and 33–35 in the Cluster C structures had a very low helical probability, while these residues are parts of helices in the 1AML structure. Furthermore, the maximum helix probability in the Cluster C structures

was 0.25. These inconsistencies are due to the flexibility of A β (12–36) in water. Comparison of the CD experiment of A β in water (Fezoui and Teplow 2002; Chen et al. 2006) with the NMR experiment of A β in TFE/water (Sticht et al. 1995) tells us that the structure of A β in water should not exactly coincide with that in TFE/water. In fact, the 20 NMR models of the 1AML structure were far from the clusters in the PCA subspace (data not shown), and the average main-chain RMSD between the Cluster C structures and the 20 NMR models was 6.9 Å (standard deviation = 1.2 Å), where the structural superposition was done for residues 12–36.

To assess in more detail the similarity between the 1AML structure and the structures in Cluster C, we analyzed the residue–residue contacts in the experimental and computational structures. First, we calculated natively contacting residue–residue pairs (NCRP) in the well-structured region (residues 12–36) of the 1AML structure. An NCRP was assigned to a residue pair when the heavy atom–heavy atom minimum distance between the residues was smaller than r_{nc} (here, we set $r_{\text{nc}} = 5.0$ Å). Then, we obtained 31 NCRPs from the 1AML structure by eliminating pairs of residues i and j with $|i - j| < 3$. Next, we

Figure 3. (A) Helix probability at each residue site from sampled conformations and from NMR structure (1AML) (Sticht et al. 1995). The helix probability was calculated with the program DSSP (see Materials and Methods). The probability from Cluster C is shown by a solid line, that from the other clusters by a broken line, and that from the 1AML structure (average of 20 NMR models) by a dotted line. (B) One of the helix-rich structures in Cluster C. (C) An NMR model from the 1AML structure.

calculated the reproduction ratio (Q value) of the NCRPs in each sampled structure. The Q value averaged over the structures in Cluster C was 0.411, and that averaged over those in other clusters was 0.238. The computed Q values again show that the structures in Cluster C have a similarity with the 1AML structure.

β -Structural Clusters A and B

Figure 4 displays the tertiary structures taken from Cluster A or B. These structures involve strand-strand H-bonds, although there is no ideal β -hairpin structure in the clusters. The strand-strand H-bonds are seen even in the structure p_1 , which involves helical residues (the numbers of helical and β -structural residues in p_1 are 7 and 4, respectively, according to the assessments by the DSSP program; Brahms and Brahms 1980). The majority of the β -rich structures in Cluster A involved seven common strand-strand H-bonds ($^{16}\text{O}-^{32}\text{HN}$, $^{18}\text{HN}-^{30}\text{O}$, $^{18}\text{O}-^{30}\text{HN}$, $^{20}\text{HN}-^{28}\text{O}$, $^{20}\text{O}-^{28}\text{HN}$, $^{22}\text{HN}-^{26}\text{O}$, and $^{22}\text{O}-^{25}\text{HN}$), although each β -rich structure does not necessarily involve all of the seven bonds. The strand-strand

H-bond patterns of some the β -rich structures exhibited one-residue sliding against those listed above. All of the β -rich structures in Cluster B involved six common strand-strand H-bonds ($^{13}\text{O}-^{36}\text{HN}$, $^{15}\text{HN}-^{34}\text{O}$, $^{17}\text{O}-^{34}\text{HN}$, $^{19}\text{HN}-^{32}\text{O}$, $^{19}\text{O}-^{32}\text{HN}$, and $^{21}\text{HN}-^{30}\text{O}$), although each β -rich structure does not necessarily involve all of the six. No other strand-strand H-bond patterns than listed above exist in Cluster B.

The free-energy barrier between Clusters A and B may be caused by the rearrangement of strand-strand H-bonds. The right panel of Figure 4 indicates the positions of residues 27–30 in yellow for each tertiary structure. In Cluster A, the yellow region is located in the C-terminal strand (see p_1 – p_3 in Fig. 4), although in Cluster B, residues 27 and 28 are located at the turn region (see p_4 and p_5 in Fig. 4). When the conformation moves from Cluster A to B, all of the strand-strand H-bonds in Cluster A should be broken to generate those in Cluster B. Besides, the β -rich structures in Cluster B (see p_4 and p_5) involve a small helix near the turn region, although those in Cluster A (see p_1 , p_2 , and p_3) do not involve such a small helix near the turn region. The

Figure 4. Representative structures in Clusters A and B. *Left* panel is distribution of sampled conformations focusing on Clusters A and B. *Right* panel displays the tertiary structures picked from the *left* panel: p_1 , p_2 , and p_3 are taken from Cluster A, and p_4 and p_5 from Cluster B. Structure 2BEG in *right* panel is from NMR structure of A β (17–40) in fibril (Luhers et al. 2005), where each NMR model consists of five chains of A β (17–40). We displayed the segment with residues 17–36 in model 1. See Fig. 2 legend for colors of dots in the *left* panel. Yellow region of tertiary structure in the *right* panel indicates the segment with residues 27–30. Broken lines represent H-bonds, and the N-terminal is colored red.

generation/annihilation of the small helix requires strand-strand H-bond rearrangements (data not shown).

Interestingly, both helix-rich and β -rich structures coexist in both Clusters A and B. Structures p_1 and p_2 are close to each other in Cluster A (Fig. 4) and involve similar strand-strand H-bond patterns. The difference between p_1 and p_2 is clear: p_1 involves a helix in the N-terminal region and its strands are smaller than those of p_2 , while p_2 involves no helical element. Thus, generation of the helix in the N-terminal region does not accompany overcoming a free-energy barrier to move from p_1 to p_2 .

Next, we compared the β -rich structures in Clusters A and B with an in-fibril A β structure determined by NMR (PDB ID = 2BEG) (Luhers et al. 2005). An essential difference between the 2BEG structure and the currently computed structures is: The intra-chain H-bonds stabilize the monomeric β -rich structures, while the interchain H-bonds stabilize the 2BEG structure in the fibril. Despite the essential difference, the Cluster B conformations are more similar to the 2BEG structure than the Cluster A conformations are, because the yellow region, which connects the two strands in the 2BEG structure (Fig. 4), is partly located at the turn region in Cluster B, although the yellow region is embedded in the C-terminal-side strand in Cluster A.

Random conformational Clusters D and E

Clusters D and E are predominantly composed of random structures. Compared to the other clusters, these rarely

include characteristic secondary structures. However, there are slight differences between Clusters D and E: Cluster D involves a few β -rich structures, and Cluster E consists of completely random structures (Fig. 2).

Here, we consider another in-fibril A β structure (not deposited to PDB) determined by an NMR experiment (Petkova et al. 2002), where an intra-polypeptide salt bridge is formed between ^{23}Asp and ^{28}Lys . In this NMR structure, the salt bridge is essentially important to form a turn-like structure between the two β -strands (residues 12–24 and 30–40). Thus, we searched the intra-polypeptide salt bridge in the sampled conformations and found that Cluster D frequently involved the salt bridge (Fig. 5) whereas Cluster E did not. The salt-bridged structures in Cluster D, however, did not involve a β -strand (see p_6 and p_7 in Fig. 5).

In this analysis, we noticed that some β -rich structures in Clusters A and B involved a salt bridge (Fig. 5). These salt-bridged β -rich structures were significantly different between Clusters A and B: Those in Cluster B involved a small helix near the salt bridge (see p_{11}), whereas those in Cluster A did not (see p_9 and p_{10}). Remember that the small helix was generally found in the β -rich structures in Cluster B (Fig. 4). Therefore, the shape of the salt-bridged β -rich structures in Cluster A is more similar with the in-fibril β -structures (Petkova et al. 2002). However, we should note again that there is a significant difference between the in-fibril β structures and the sampled salt-bridged β -rich structures: The in-fibril

Figure 5. PCA conformational subspace displayed to indicate positions (blue dots) of structures with salt bridge ^{23}Asp – ^{28}Lys . The subspace is constructed by the first, second, and third eigenvectors, although the axes are not shown. The salt bridge was assigned when the distance between the side-chain oxygen atoms of ^{23}Asp and the side-chain nitrogen atom of ^{28}Lys was <3.0 Å. Tertiary structures p_6 – p_{11} are picked from the subspace: The structures p_6 , p_7 , p_9 – p_{11} involve the salt bridge, and p_8 is a conformation randomly picked from Cluster E. Yellow-colored residues are ^{23}Asp and ^{28}Lys . The N-terminal is colored red.

structure is stabilized by inter-polypeptide H-bonds, whereas the computational structure is stabilized by intra-polypeptide H-bonds.

Analysis of an NMR structure, 1HZ3, determined in water

In contrast to the CD-experiment reports that A β is disordered in water (Fezoui and Teplov 2002; Chen et al. 2006), an NMR experiment (Zhang et al. 2000) proposed a stable structure of A β (10–35) in water, where a compact core region consisting of residues 16–27 is formed (PDB ID = 1HZ3; 15 models deposited) and does not visually involve either an α or a β secondary-structure element. It is interesting to analyze if the currently sampled conformations have some similarity with this 1HZ3 structure. A segment consisting of residues 12–35 is the common region for the 1HZ3 structure and our system A β (12–36). We reconstructed a PCA subspace for this common region. Figure 6 displays the conformational distribution at 300°K in the subspace together with the 15 NMR models (magenta dots in Fig. 6) of the 1HZ3 structure. The tertiary structure of model 12 in the 1HZ3 structure is also displayed in Figure 6, where the

1HZ3 structure is embedded in Cluster C. Remember that Cluster C, which is the helix-rich cluster, had a similarity with the stable helical structure (i.e., 1AML structure) in 40% TFE/water. A tertiary structure (p_{12}) near the NMR models is shown in Figure 6.

In a precise sense, the PCA subspace reconstructed for A β (12–35) is different from that constructed for A β (12–36). However, the clusters in Figure 5 are very similar with those in Figure 6 because most of the residue–residue distances used for constructing the PCA subspace are exactly the same between the two segments: The addition of only one residue (residue 36) to A β (12–35) generates A β (12–36). Note that the orientations to view the PCA subspace are different in Figures 5 and 6.

Next, we compared the nuclear Overhauser effect (NOE) spectra used for determining the 1HZ3 structure with the currently sampled conformations. There are 28 long-range NOE pairs in the common region, A β (12–35). A long-range pair was defined so that residues i and j , to which the NOE restraints are assigned, satisfy a condition of $|i - j| \geq 3$. The long-range NOEs are related to the global fold, and short-range ones ($|i - j| < 3$) to the local organization of the polypeptide. To assess how

Figure 6. Conformational distribution of sampled conformations at 300°K and 15 NMR models of the 1HZ3 structure. PCA was done with computing the minimum residue–residue distances in the common region (residues 12–35). The first, second, and third eigenvectors were used to construct the PCA space (the axes are not shown). Magenta dots represent the 15 NMR models. Sampled structures satisfying 14 or more restraints among 28 long-range NOE restraints are dark blue, and those satisfying 9–13 restraints are sky blue. The largest number of the satisfied NOE restraints was 17. The structure of p_{12} is an example of conformations near the NMR models; that of model 12 of the 1HZ3 structure is displayed; structures of p_{13} and p_{14} are those satisfying more than 14 long-range NOE restraints.

the long-range NOE restraints are satisfied in the sampled conformations, we calculated $R = \langle 1/r^6 \rangle^{-1/6}$, where r is the atom–atom distance for the NOE pair, and $\langle \dots \rangle$ represents the ensemble average over the 841 conformations sampled at 300°K. Figure 7 demonstrates the obtained R for the long-range pairs together with their upper bounds, R_{NMR} , determined by the NMR experiment (Zhang et al. 2000). We found that 15 pairs (54% = 15/28) satisfied the NOE restraints ($R \leq R_{\text{NMR}}$) and 8 pairs (29%) weakly violated the upper bounds as $R - R_{\text{NMR}} \leq 1 \text{ \AA}$. Two pairs (7%) significantly violated the upper bounds as $1 \text{ \AA} < R - R_{\text{NMR}} \leq 2 \text{ \AA}$, and three pairs (11%) strongly violated the upper bounds as $R - R_{\text{NMR}} > 2 \text{ \AA}$. These percentages for the satisfied long-range restraints are comparable with those obtained from a replica-exchange MD simulation of A β (10–35) by Shea's group (Baumketner and Shea 2007): 50% for $R \leq R_{\text{NMR}}$, 20% for $R - R_{\text{NMR}} \leq 1 \text{ \AA}$, and 17% for $1 \text{ \AA} < R - R_{\text{NMR}} \leq 2 \text{ \AA}$.

Figure 6 displays the positions of structures satisfying the long-range NOE restraints relatively well in the PCA subspace. Interestingly, these conformations were located not only in Cluster C but also in Cluster A. Furthermore,

those conformations in Cluster A adopted β -rich structures (see p_{13} and p_{14} in Fig. 6).

Figure 8 shows the relation between the number of satisfied long-range NOE restraints (N_{satis}) and a structural similarity (main-chain RMSD) to the 1HZ3 structure. One may expect that the larger the N_{satis} , the smaller the main-chain RMSD. However, Figure 8 does not show such a correlation. Figure 9 indicates the relation between N_{satis} and the number of helical residues for the sampled structures. Although the standard deviation is large, Figure 9 shows that the greater the value of N_{satis} , the higher the number of helical residues for $N_{\text{satis}} \leq 12$. In contrast, the structures for $N_{\text{satis}} \geq 13$ had small numbers of helical residues, which may suggest that the simultaneous satisfaction of many NOE restraints causes some structural frustrations in a helical structure or may be due to a statistical error by a small number (25, which is 3.0% of the 841 structures) of sampled structures for $N_{\text{satis}} \geq 13$.

Figure 9 suggests that the 1HZ3 structure might be helix-related. Then, we compared the 1HZ3 structure with various structures of fold classes (all- α , all- β , α/β , and $\alpha + \beta$ classes) taken from the database. After discarding

Figure 7. Distance R ($= \langle 1/r^6 \rangle^{-1/6}$) for each long-range NOE pair together with the upper bound determined in the NMR experiment (Zhang et al. 2000). Solid line represents R averaged over 841 sampled structures at 300°K; broken line represents the upper bounds from the NMR experiment.

the highly disordered terminal regions (residues 10–15 and 28–35) of the 1HZ3 structure, we calculated the NCRPs for the remaining 12-residue core regions of the 1HZ3 structure. Next, a set of 510 protein folds was taken from the SCOP database (release 1.65; Murzin et al. 1995) for the four classes: 133 folds from the all- α class, 100 folds from the all- β class, 83 folds from the α/β class, and 194 folds for the $\alpha + \beta$ class. The number 510 was a result of eliminating folds shorter than 50 and or longer than 450 residues and by removing multichain folds. We randomly picked a protein from each fold, then randomly picked a 12-residue segment from each of the picked proteins, and computed the residue–residue contacts for each of the 510 segments. Lastly, we calculated the Q value of each segment against the NCRPs of the core region of the 1HZ3 structure. In this computation we changed the distance r_{nc} to obtain the NCRPs. The ratio averaged over the segments in each class is shown in Figure 10. It is clear that the highest structural similarity is assigned to the all- α class and the lowest to the all- β class. The similarities for the α/β and $\alpha + \beta$ classes are the intermediates. Since the segments were randomly taken from the database, those for the all- α class did not necessarily involve an ideal helix constructed by the 12 residues. The reason for the random choice is to give a structural variety to the segments from the all- α class. Figure 10 indicates that the core region of the 1HZ3 structure has the highest similarity with helical structures

than the other structures. We have done the same analysis with changing the selection of 510 folds and obtained a very similar result.

Discussion

The CD study (Fezoui and Teplov 2002) provides the secondary-structure contents for the monomeric full-length

Figure 8. Number (N_{satis}) of satisfied long-range NOE restraints vs. main-chain RMSD of sampled structures at 300°K against model 1 of the 1HZ3 structure. Error bars show the standard deviations in each ensemble at given N_{satis} .

Figure 9. Number (N_{satis}) of satisfied long-range NOE restraints vs. number of helical residues. The number of helical residues is an average over structures at each N_{satis} . The program DSSP (Brahms and Brahms 1980) was used to judge the helicity of each residue. Error bars show the standard deviations.

A β (1–40) in water. Our simulation reproduced secondary-structure contents agreeing with those experimentally measured values. This agreement ensured us to proceed to atomic-level analysis of the sampled structures.

Here, we consider the first question addressed earlier. The conformational ensemble at 300°K consists of five clusters. There is no major cluster in the ensemble, and a large structural variety exists in the sampled conformations. Thus, the structure of A β (12–36) in water should be categorized as the disordered state. However, each cluster has peculiar structural characteristics of its own. Thus, the disordered state of A β is realized not by a plain but by a rugged free-energy landscape consisting of secondary-structure specific clusters. The experimentally measured secondary-structure contents are the consequence of elemental contributions from the clusters.

Next, we consider the second question. The structures in Cluster C have similarity with the 1AML structure. This similarity was weak when we used the RMSD values for the measure of similarity. However, the similarity became clearer when we used the positions of helices on the sequence and the native-contact patterns for the measure. The CD experiment of A β (Fezoui and Teplow 2002; Chen et al. 2006) has shown that the helix content gradually increases with increasing the TFE concentration. Thus, we speculate that Cluster C might mature to the free-energy global minimum when the concentration of TFE is increased, i.e., when the two unstable helices in Cluster C become stable.

Now we consider the third question. The β -structure content is mainly contributed by Clusters A and B. Interestingly, these clusters are composed of different types of β -rich structures, and there is a free-energy

barrier between them. This free-energy barrier is probably due to the rearrangements of strand–strand H-bonds. One may ask if there is any relation between the computed β -rich structures and the in-fibril structures. We treated the monomeric state of A β (12–36) in the current study. Thus, we should not directly relate the monomeric free-energy landscape with fibril formation. However, it is interesting to check structural similarities between the currently sampled structures and the experimentally determined fibril structures (Petkova et al. 2002; Sciarretta et al. 2005). A study has reported that A β fibrils exhibit multiple distinct morphologies (Petkova et al. 2005). In other words, various types of fibrils and tertiary structures of A β are possible. As shown in the Results, the monomeric A β (12–36) can fold into two types of hairpin clusters (Clusters A and B), and there are some superficial similarities between the NMR structures in fibril and the currently sampled structures. However, there was an essential difference between the NMR and computed structures: The experimental β -structures were stabilized by the inter-polypeptide strand–strand H-bonds in fibril. In contrast, the computed β -rich structures were stabilized by intra-polypeptide ones. We presume that in a multichain system, a large free-energy barrier is required to rearrange the intra-polypeptide H-bonds toward the inter-polypeptide H-bonds, because all of the intra-polypeptide H-bond ones should be broken to convert A β to the in-fibril form.

We should note here the existence of the salt bridge ^{23}Asp – ^{28}Lys in the sampled conformations. An experiment (Sciarretta et al. 2005) has reported that this salt-bridge formation essentially controls the aggregation rate of A β . Baumketner and Shea have shown that this salt bridge is easily formable in the random structures at room temperature in explicit water (Baumketner and Shea

Figure 10. Similarity (Q value) of NCRP patterns in the 1HZ3 structure with residue–residue contacts in segments taken from the SCOP database (Murzin et al. 1995). The X-axis represents the value of residue–residue contact distance (r_{nc}) used to calculate NCRPs.

2007). The current study showed that the salt bridge is formable in either the β -rich structures or the random structures (Fig. 5). These results led us to focus on how the salt bridge is formed in a multichain system in a future work.

We failed to obtain structures of a small RMSD to the 1HZ3 structure. Here, we list some inconsistencies between the currently sampled and the 1HZ3 structures: Figure 8 clearly indicates that the conformations satisfying the long-range NOE restraints do not result in a small RMSD. Despite the 1HZ3 structure being embedded in Cluster C (helix-rich cluster), the sampled structures (blue dots in Fig. 6) that most satisfied the long-range NOE restraints were embedded in Cluster A. Analysis of NCRP patterns showed that the 1HZ3 structure provided the highest similarity with all- α class proteins, in spite of 1HZ3 involving neither α nor β secondary-structure elements. We cannot explain these inconsistencies only from the current simulations. However, we note that an inconsistency also exists among the experimental studies. The 1HZ3 structure has a well-structured coil region (residues 16–27), where no secondary-structure element is involved. Contrarily, the CD experiment (Fezoui and Teplov 2002; Chen et al. 2006) has judged that A β is disordered in water, and that the disordered state involves 25% β and 5% α contents. If one considers that this inconsistency should be solved experimentally, we cannot introduce any proposal for the inconsistency. Instead, we merely state that our simulation supports the conclusion from the CD experiments.

Lastly, we summarize the main results from the current study: The free-energy landscape of A β (12–36) in water was not plain but rugged with five conformational clusters, and the secondary-structure contents average over the clusters agreed well with those from the CD experiment. Each cluster had particular structural characteristics (two β -structural clusters, one helical-structural cluster, and two random-structural clusters). The helical cluster had structural similarity with the tertiary structure (1AML structure) that was determined in a 40% TFE/water solution. There were two types of β -structural clusters between which a free-energy barrier existed.

Materials and Methods

The CD experiment reported that A β (1–40) adopts disordered structures in water (Fezoui and Teplov 2002; Chen et al. 2006). On the other hand, with 40% (v/v) TFE/water, an NMR study (Sticht et al. 1995) reported that A β (1–40) composed of two stable helices, helix 1 (residues 12–22) and helix 2 (residues 31–36), and a coil region (residues 23–30) links the helices (PDB ID = 1AML; 20 NMR models deposited). The terminal regions (residues 1–11 and 37–40) are unstructured in the NMR models. The primary aim of the current study is to make the connection between the disordered structures in water and the stable helical

structure in 40% TFE/water. Thus, the unstructured terminal regions in either solution are beyond the scope of the current study, and we removed the regions before the simulation to reduce the computational tasks. The computed segment A β (12–36) is the most important region for studying the structure of A β in either water or TFE/water. The amino acid sequence of A β (12–36) is [Ace-¹²Val-¹³His-¹⁴His-¹⁵Gln-¹⁶Lys-¹⁷Leu-¹⁸Val-¹⁹Phe-²⁰Phe-²¹Ala-²²Glu-²³Asp-²⁴Val-²⁵Gly-²⁶Ser-²⁷Asn-²⁸Lys-²⁹Gly-³⁰Ala-³¹Ile-³²Ile-³³Gly-³⁴Leu-³⁵Met-³⁶Val-Nme], where Ace and Nme are the N-terminal acetyl and C-terminal *N*-methyl groups, respectively, introduced to cap the termini.

We set up the simulation system as follows. A sphere (sphere 1; diameter = 47 Å) consisting of water molecules was prepared. The water molecules in sphere 1 had been configurationally equilibrated at a density of 1 g/cc at 300°K in advance. Then, A β (12–36) was immersed in sphere 1 with the geometrical center of A β (12–36) set at the center of sphere 1. Water molecules overlapping polypeptide atoms were removed. The system consisted of 5305 atoms (397 polypeptide atoms and 1636 water molecules).

Force-field parameters for the polypeptide and water molecule were taken from an AMBER-based hybrid force field (Kamiya et al. 2005) and the flexible TIP3P water model (Jorgensen et al. 1987), respectively. The AMBER-based hybrid force field (E_{hybrid}) is a mixture of AMBER parm94 (E_{94}) (Cornell et al. 1995) and parm96 (E_{96}) (Kollman et al. 1997): $E_{\text{hybrid}}(w) = (1 - w)E_{94} + wE_{96}$, where w is a weighting factor. The difference between E_{94} and E_{96} exists only in the main-chain torsion angle energy terms. Thus, E_{hybrid} was actually computed by simply hybridizing the torsion energy terms of E_{94} and E_{96} . A quantum-chemical calculation has shown that E_{hybrid} is better than parm96 and parm94 for $0.35 < w < 1.0$, and a McMD simulation has shown that E_{hybrid} provides different stable structures depending on the amino acid sequence (Kamiya et al. 2005). We examined eight intra-polypeptide force fields, setting w to 0.50, 0.60, 0.65, 0.70, 0.75, 0.85, 0.90, and 1.0, respectively.

A computer program, PRESTO, ver. 3 (Morikami et al. 1992), was used for the McMD simulation with the following simulation conditions: time step, 1 fs; the SHAKE method (Ryckaert et al. 1977) to constrain covalent bonds between heavy atoms and hydrogen atoms; cell multipole expansion method (Ding et al. 1992) to compute electrostatic interactions; and temperature control by a constant-temperature method (Evans and Morriss 1983). To avoid evaporation, a harmonic potential was applied to water-oxygen atoms only when they were outside sphere 1. To keep A β (12–36) at the sphere center, the momentum and angular momentum of A β (12–36) were constrained to zero during the simulation. To avoid hydrophobic amino acid residues being exposed to outside sphere 1, another harmonic potential was applied to the polypeptide heavy atoms only when the atoms were outside a smaller sphere (diameter = 41 Å), which was centered at the center of sphere 1.

The McMD simulation method (Nakajima et al. 1997) is explained briefly below. A modified potential energy is defined as $E_{\text{McMD}} = E + RT_0 \ln[P(E, T_0)]$, where E is the original potential energy of the system, T_0 the simulation temperature, and $P(E, T_0)$ the canonical energy distribution at T_0 . The temperature T_0 is usually set to a high value (700°K in this work) so that the conformation can overcome high potential-energy barriers. E_{McMD} is used to calculate forces acting on atoms: force = $-\nabla E_{\text{McMD}}$. If $P(E, T_0)$ is accurately estimated in a range of potential energy, the McMD simulation provides a flat energy distribution in the range. Since $P(E, T_0)$ is unknown a priori,

iterative runs are required through which $P(E, T_0)$ gradually converges to the flat energy distribution. Conformations sampled in the last run (i.e., the productive run) of the iteration are used for the conformational analysis. The canonical distribution $P(E, T)$ at any temperature T in the energy range is derived from $P(E, T_0)$ with the reweighting technique (Higo et al. 2001; Kamiya et al. 2002; Ikeda and Higo 2003; Ikeda et al. 2003) as follows: The density of states $n(E)$ is related to $P(E, T_0)$ as $P(E, T_0) = n(E) \exp[-E/RT_0]/Z(T_0)$, where $Z(T_0)$ is a partition function at T_0 . $P(E, T)$ is then derived as $P(E, T) = n(E) \exp[-E/RT]/Z(T) = P(E, T_0) \exp[E/RT_0] \exp[-E/RT]Z(T_0)/Z(T)$. The term $Z(T_0)/Z(T)$ can be regarded as a normalization factor to calculate $P(E, T)$. Choosing conformations with a statistical weight of $P(E, T)$ from all of the sampled conformations, we can generate a thermodynamically equilibrated conformational ensemble at T . The starting conformation of A β (12–36) for the first run was a random conformation, and the starting one for the i th run was the last conformation of the $(i - 1)$ th run. It took 8 wk to simulate a force field using 12 Pentium-4 (3.2-GHz) machines. Thus, the entire simulations required 64 wk.

To analyze the free-energy landscape, PCA was applied on the obtained conformational ensemble, as follows. The minimum heavy atom–heavy atom distance (d_{ij}) between residues i and j was calculated for each sampled conformation (number of residue pairs = 276 with condition of $|i - j| \geq 3$). Then, every conformation was represented by a vector as $\mathbf{q} = [q_1, q_2, \dots, q_{276}]$, where $q_1 = d_{12,15}$, $q_2 = d_{12,16}$, $\dots$, $q_{276} = d_{33,36}$. Next, a variance–covariance matrix, C , was calculated: $C_{mn} = \langle q_m q_n \rangle_T - \langle q_m \rangle_T \langle q_n \rangle_T$, where C_{mn} is the (m, n) th element, and $\langle \dots \rangle_T$ represents the ensemble average over conformations at T . By diagonalizing C , a set of eigenvectors and eigenvalues were obtained: the k th eigenvector and eigenvalue are denoted as $\mathbf{v}_k(T)$ and $\lambda_k(T)$, respectively. The eigenvectors satisfy the equation $\mathbf{v}_i(T) \cdot \mathbf{v}_j(T) = \delta_{ij}$. We arranged the eigenvalues as $\lambda_i > \lambda_j$ when $i < j$. The eigenvectors construct a high-dimensional conformational space (PCA space) as coordinate axes. Then, we constructed a three-dimensional (3D) subspace by the eigenvectors, $\mathbf{v}_1(T)$, $\mathbf{v}_2(T)$, and $\mathbf{v}_3(T)$, most contributing to the conformational distribution of the sampled conformations in the PCA space. A conformation, $\mathbf{q}'$, at T was then projected on the subspace as $\mathbf{c}_i = \mathbf{v}_i(T) \cdot (\mathbf{q}' - \langle \mathbf{q} \rangle_T)$. The coordinates $[c_1, c_2, c_3]$ represent the position of $\mathbf{q}'$ in the 3D subspace. Projecting all of the conformations sampled at T in the 3D subspace, we obtained the conformational distribution. This distribution can be regarded as an image of a free-energy landscape.

To compute a secondary-structure content of the sampled conformations, we used the program DSSP (Brahms and Brahms 1980). This program checks the H-bond patterns from the atomic coordinates of a polypeptide and assigns the secondary structure to every amino acid residue. We eliminated the N- and C-termini (i.e., Ace and Nme) to compute the secondary-structure content for each sampled conformation. We classified amino acid residues in α -helix, 3_{10} -helix, and π -helix as helical residues. The β -structural residues consist of those in isolated β -bridge, extended strand, and β -turn. Amino acid residues in other states were classified in random structural residues. The helix content of a conformation was simply defined as $P_{\text{helix}} = N_{\text{helix}}/N_{\text{all}}$, where N_{helix} is the number of helical residues in the conformation, and N_{all} the total number of residues. Then, the helix content of the conformational ensemble at 300°K is the average of P_{helix} over the conformations in the ensemble. The β -structural content and random structural content were calculated similarly.

Acknowledgments

We are grateful to Dr. K. Yamaguchi, Center for Emerging Infectious Diseases, Gifu University, and Dr. Joan-Emma Shea, Department of Chemistry and Biochemistry, University of California, for helpful discussions.

References

- Baumketner, A. and 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.
- Brahms, S. and Brahms, J. 1980. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.* **138**: 149–178.
- Chen, Y.R., Huang, H.B., Chyan, C.L., Shiao, M.S., Lin, T.H., and Chen, Y.C. 2006. The effect of A β conformation on the metal affinity and aggregation mechanism studied by circular dichroism spectroscopy. *J. Biochem.* **139**: 733–740.
- Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**: 5179–5197.
- Ding, H.-Q., Karasawa, N., and Goddard III, W.A. 1992. Atomic level simulations on a million particles: The cell multipole method for Coulomb and London nonbond interactions. *J. Chem. Phys.* **97**: 4309–4315.
- Evans, D.J. and Morriss, G.P. 1983. The isothermal/isobaric molecular dynamics ensemble. *Phys. Lett. A* **98**: 433–436.
- Fezoui, Y. and Teplow, D.B. 2002. Kinetic studies of amyloid β -protein fibril assembly. *J. Biol. Chem.* **277**: 36948–36954.
- Higo, J., Ito, N., Kuroda, M., Ono, S., Nakajima, N., and Nakamura, H. 2001. Energy landscape of a peptide consisting of α -helix, 3_{10} -helix, β -turn, β -hairpin, and other disordered conformations. *Protein Sci.* **10**: 1160–1171.
- Ikeda, K. and Higo, J. 2003. Free-energy landscape of a chameleon sequence in explicit water and its inherent α/β bifacial property. *Protein Sci.* **12**: 2542–2548.
- Ikeda, K., Galzitskaya, O.V., Nakamura, H., and Higo, J. 2003. β -Hairpins, α -helices, and the intermediates among the secondary structures in the energy landscape of a peptide from a distal β -hairpin of SH3 domain. *J. Comput. Chem.* **24**: 310–318.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impley, R.W., and Klein, M.L. 1987. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**: 926–935.
- Kamiya, N., Higo, J., and Nakamura, H. 2002. Conformational transition states of a β -hairpin peptide between the ordered and disordered conformations in explicit water. *Protein Sci.* **11**: 2297–2307.
- Kamiya, N., Watanabe, Y.S., Ono, S., and Higo, J. 2005. AMBER-based hybrid force field for conformational sampling of polypeptides. *Chem. Phys. Lett.* **401**: 312–317.
- Kollman, P.A., Dixon, R.W., Cornell, W.D., Chipot, C., and Pohorille, A. 1997. The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In *Computer simulations of biological systems* (eds. W.F. van Gunsteren et al.), pp. 83–96. Elsevier, New York.
- Luhers, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Dobeli, H., Schubert, D., and Riek, R. 2005. 3D structure of Alzheimer's amyloid- β (1–42) fibrils. *Proc. Natl. Acad. Sci.* **102**: 17342–17347.
- Morikami, K., Nakai, T., Kidera, A., Saito, M., and Nakamura, H. 1992. PRESTO: A vectorized molecular mechanics program for biopolymers. *Comput. Chem.* **16**: 243–248.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**: 536–540.
- Nakajima, N., Nakamura, H., and Kidera, A. 1997. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J. Phys. Chem. B* **101**: 817–824.
- Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F., and Tycko, R. 2002. A structural model for Alzheimer's β -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci.* **99**: 16742–16747.

- Petkova, A.T., Leapman, R.D., Guo, Z., Yau, W.N., Mattson, M.P., and Tycko, R. 2005. Self-propagating, molecular-level polymorphism in Alzheimer's β -amyloid fibrils. *Science* **307**: 262–265.
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H.J.C. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**: 327–341.
- Sciarretta, K.L., Gordon, D.J., Petkova, A.T., Tycko, R., and Meredith, S.C. 2005. A β 40-Lactam(D23/K28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry* **44**: 6003–6014.
- Sticht, H., Bayer, P., Willbold, D., Dames, S., Hilbich, C., Beyreuther, K., Frank, R.W., and Rosch, P. 1995. Structure of amyloid A4-(1–40)-peptide of Alzheimer's disease. *Eur. J. Biochem.* **233**: 293–298.
- Sugita, Y. and Okamoto, Y. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**: 141–151.
- Wei, G. and Shea, J.E. 2006. Effects of solvent on the structure of the Alzheimer amyloid- β (25–35) peptide. *Biophys. J.* **91**: 1638–1647.
- Zhang, S., Iwata, K., Lachenmann, M.J., Peng, J.W., Li, S., Stimson, E.R., Lu, Y., Felix, A.M., Maggio, J.E., and Lee, J.P. 2000. The Alzheimer's peptide A β adopts a collapsed coil structure in water. *J. Struct. Biol.* **130**: 130–141.