# The conformations of the amyloid-$\beta$ (21–30) fragment can be described by three families in solution

Wei Chen
*Département de Physique, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada and Regroupement Québécois sur les Matériaux de Pointe, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada*

Normand Mousseau[a)]
*Département de Physique, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada; Regroupement Québécois sur les Matériaux de Pointe, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada; and Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique et Université Paris 7, 13 Rue Pierre et Marie Curie, 75005 Paris, France*

Philippe Derreumaux
*Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique et Université Paris 7, 13 Rue Pierre et Marie Curie, 75005 Paris, France*

(Received 12 June 2006; accepted 25 July 2006; published online 29 August 2006)

Alzheimer's disease has been linked to the self-assembly of the amyloid-$\beta$ protein of 40 and 42 residues. Although monomers are in equilibrium with higher-order species ranging from dimers to heptamers, structural knowledge of the monomeric amyloid-$\beta$ (A$\beta$) peptides is an important issue. Recent experimental data have shown that the fragment (21–30) is protease-resistant within full-length A$\beta$ peptides and displays two structural families in solution. Because the details of the A$\beta_{21-30}$ structures found using distinct force fields and protocols differ at various degrees from those of the NMR structures, we revisit the conformational space of this peptide using the activation-relaxation technique (ART nouveau) coupled with a coarse-grained force field (OPEP v.3.0). We find that although A$\beta_{21-30}$ does not have a secondary structure, it dominantly populates three structural families, with a loop spanning residues Val24-Lys28. The first two families, which differ in the nature of the electrostatic interactions, satisfy the five interproton rotating frame nuclear Overhauser effect spectroscopy (ROESY) distances and superpose well onto the NMR structures. The third family, which cannot be seen by ROESY NMR experiments, displays a more open structure. This numeric study complements the experimental results by providing a much more detailed description of the dominant structures. Moreover, it provides further evidence of the capability of ART OPEP in providing a reliable conformational picture of peptides in solution.
 [DOI: 10.1063/1.2337628]

## I. INTRODUCTION

The neuropathogenesis of Alzheimer's disease (AD) has been associated with amyloid-$\beta$(A$\beta$)-protein fibrils. However, the exact process by which these fibrils cause AD is still unclear. Over the last few years, it has been discovered that aggregates prior to amyloid deposition (or oligomers) are also neurotoxic, suggesting that these transient assemblies could play a major role in AD.[1–3] This finding shifts the focus onto understanding the early steps of aggregation in order to develop protocols that could prevent the formation of these structures or destabilize the already formed oligomers.

A$\beta$ is present in two dominant forms with 40 (A$\beta_{1-40}$) and 42 (A$\beta_{1-42}$) amino acids. Differing by only two residues at the C-terminus, Ile41 and Ala42, the oligomerization pathways and the end products of these two forms are different,[4] providing researchers with a natural tool to understand the driving forces for aggregation. Polymerization starts with the formation of pentameric/hexameric species for A$\beta_{1-42}$ versus monomeric, dimeric, trimeric, and tetrameric species in equilibrium for A$\beta_{1-40}$.[4,5] Both A$\beta$ models share a cross-$\beta$ structure with parallel $\beta$ sheets, but residues 1–8 are disordered in A$\beta_{40}$ versus residues 1–15 in A$\beta_{42}$, and the salt bridge between residues Asp23 and Lys28 is intramolecular in A$\beta_{40}$, but intermolecular in A$\beta_{42}$.[6,7]

To facilitate the understanding of the aggregation mechanisms, researchers turned to fragments of A$\beta$, which also form amyloid structures by themselves. These fragments span the residues 16–22,[8,9] 11–25,[10,11] 10–35,[12,13] 15–36,[14] and 1–28,[15] among others. Although these studies provide valuable information, extrapolation to full-length A$\beta$ peptides is not an easy task. Recently, Lazo *et al.* coupled limited proteolysis and liquid chromatography/mass spectrometry experiments to probe the structure within monomeric A$\beta_{1-40}$ and A$\beta_{1-42}$ peptides.[16] They identified the 21–30 region as protease resistant in the A$\beta_{1-40}$ and A$\beta_{1-42}$ peptides. By combining this finding with the propensity of the 21–30

[a)]Electronic mail: normand.mousseau@umontreal.ca

 

region for turn- or bendlike structures in the $A\beta_{1-40}$ and $A\beta_{1-42}$ peptides by solution NMR,[17] they suggested that the region (21–30) could nucleate the folding of full-length $A\beta$. In addition, using solution NMR spectroscopy, they also found that $A\beta_{21-30}$ displays two structural families characterized by a turn formed by residues Val24-Lys28.[16]

Following this report, three series of simulations have been performed on the monomer of $A\beta_{21-30}$. These include standard all-atom explicit solvent molecular dynamics,[18] replica-exchange all-atom implicit solvent molecular dynamics (REMD),[19] and discontinuous implicit solvent molecular dynamics (DMD) with a coarse-grained chain representation.[20] All simulations confirm the existence of a turn between Val24 and Lys28 stabilized by interactions between the negatively charged Glu22 (or Asp23) and the positively charged Lys28. However, the details of the populated structures found using distinct force fields and simulation methods differ from the experimentally deduced structural families.

Yet, reproducing the NMR-derived structures of the fragment $A\beta_{21-30}$ is important for three reasons. First, an engineered lactam bridge between Asp23 and Lys28 increases the $A\beta_{1-40}$ fibrillogenesis rate by three orders of magnitude, suggesting that forming a bent structure within residues 23–29 could be the rate-limiting step in $A\beta$ fibril formation.[21] Second, most disease-causing mutations are located at positions 22 and 23. Finally, because of the quality of experimental data, $A\beta_{21-30}$ can also serve as a benchmark for force fields and simulation methods that will be used for investigating the folding or assembly of full-length $A\beta$ peptides.

In this study, we use the activation-relaxation technique.[22–24] (ART nouveau) with the OPEP force field (Version 3.0).[25] ART OPEP simulations have been carried out on a number of protein models including peptides in solution ($\alpha$-helix, $\beta$-hairpin) and multimers of amyloid-forming peptides in settings varying from dimer to heptamer.[26–29] Comparison with experiments shows good accuracy in reproducing the structures of peptides in solution, for example, the difference between the experimental and simulated structures is within 1–2 Å for a $\beta$-hairpin and a zinc finger model.[30]

Here, we consider two 10 000-event ART OPEP trajectories to sample the conformational space of $A\beta_{21-30}$. Consistent with previous simulations, we find a turn spanning residues Val24-Lys28 stabilized by an attractive interaction between Glu22 or Asp23 and Lys28. These structures can be further classified into three dominant superclusters which offer an excellent agreement to date with experiments: two of these superclusters recover the families derived from NMR experiments, showing a perfect match with the measured rotating frame nuclear Overhauser effects (ROEs), while the third supercluster adds a second turn, between Glu22 and Val24, leading to an open structure, still respecting most of the ROEs.

## II. MATERIALS AND METHODS

### A. ART OPEP simulations in implicit solvent

The activation-relaxation technique explores the space of conformations by moving the system from minimum to minimum going through a common first-order transition point.[22] ART nouveau, a more recent version of the algorithm[23,24] has been applied with success to study the folding mechanisms of a $\beta$-hairpin[31] and the aggregation processes of dimers to octamers.[26–29,32]

One ART nouveau event consists of four steps. Starting from a local minimum, the atoms are moved in a randomly chosen direction until the direction of lowest local curvature of the energy landscape, given by the Hessian matrix, becomes negative. At this point, the system is then pushed against the force along the direction of negative curvature while the energy is minimized in the hyperplane perpendicular to this direction. This second step is iterated until the total force becomes smaller than a given threshold, indicating that the system has reached a transition point. Finally, the system is pushed slightly over the saddle point and is relaxed into a new local minimum, using standard minimization techniques. This move from a local minimum to a nearby one is called an event.

Events are accepted using a Metropolis criterion:[33] $P_{\text{accept}}=\min(1,e^{-\Delta E/K_BT})$, where $\Delta E$ is the difference between the new and previous energies. Because thermal vibration effects are not included in ART, the Metropolis temperature in the accept/reject criterion does not correspond to a real temperature; it is generally chosen to ensure an efficient sampling of the conformational space. Moreover, because the exact bias associated with the selection of a transition point from a random direction is ill-defined, the algorithm does not respect detailed balance.

ART is coupled to the OPEP energy function which uses a reduced off-lattice protein representation.[25,30,34,35] All amino acids are represented by their N, H, $C_\alpha$, C, and O atoms and each side chain is modeled by one sphere with an appropriate van der Waals radius and geometry with respect to the main chain and hydrophobic/hydrophilic character.[25,30,34,35] The OPEP energy function is expressed by a linear combination of four interaction types: (1) harmonic potentials for maintaining the bond lengths and bond angles of all particles, the improper dihedral angles of the side-chains and peptide bonds near equilibrium values, and $\phi$ and $\psi$ potentials for generating Ramachandran plot of reduced protein structures in agreement with all-atom protein structures, (2) backbone two-body and four-body (cooperative) hydrogen bonding interactions, (3) side-chain–side-chain interactions represented by a 12-6 potential if the side chains are hydrophobic or oppositely charged and by a six potential, otherwise, and (4) excluded-volume potentials between all particles. In this work, we use a new parameter set (OPEP Version 3.0) refined on homemade and publicly available decoys of a total of 30 proteins and peptides.[25]

### B. Details of the simulations and structural analysis

The amino acid sequence of $A\beta_{21-30}$ is AEDVGSNKGA. Two 10 000-event independent simulations starting from a fully stretched conformation are performed at a Metropolis temperature of 900 K using ART OPEP. This temperature is selected to ensure a wide sampling of the confor-

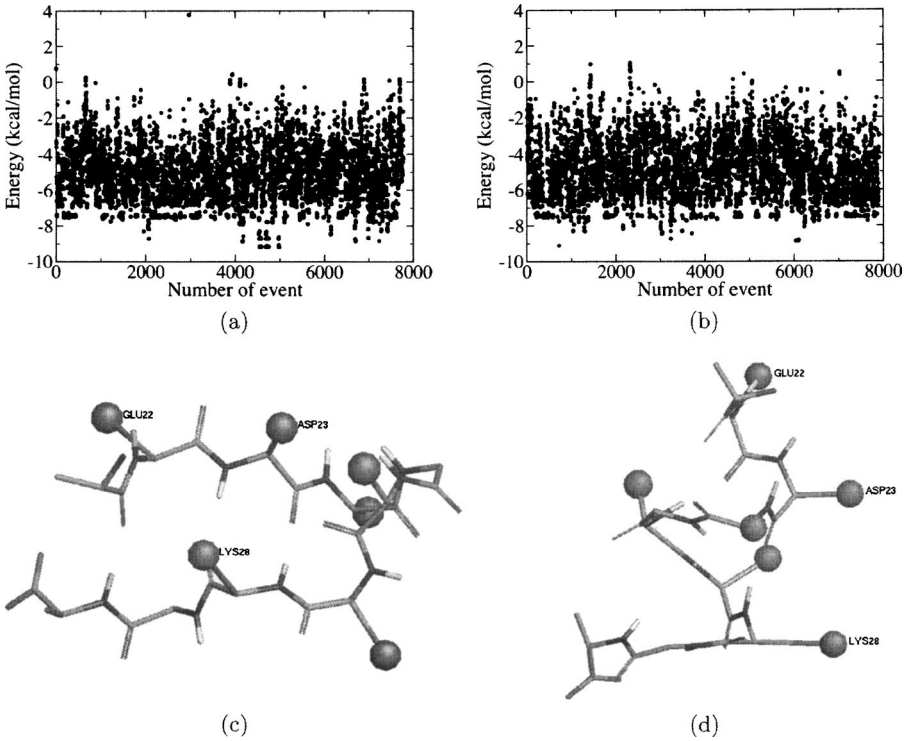


FIG. 1. Top: energy as a function of accepted event numbers in ART OPEP runs at a Metropolis temperature of 900 K for (a) run 1 and (b) run 2. Bottom: lowest-energy structure for (c) run 1 and (d) run 2. The backbone C, N, H, and O atoms are represented as sticks; the side chains of Glu22, Asp23, Val24, Ser26, Asn27, and Lys28 are represented by gray balls.

mational space and leads to an acceptation rate for these two runs of 78%.

Figure 1 shows the energy as a function of the accepted events for both simulations and Fig. 2 gives the evolution of the radius of gyration $R_g$ and end-to-end distance measured from the two $C_\alpha$ atoms of Ala21 and Ala30. From these graphics, we see that both runs explore extensively the energy spectrum—ranging from −9.2 to 3.8 kcal/mol in run 1 (from −9.1 to 1.0 kcal/mol in run 2)—and the conformational space: the end-to-end distance varies from 5.0 to 26 Å whereas $R_g$ oscillates between 4.5 and 8.2 Å.

While ART OPEP simulations sample conformations with a wide range of energies, we are interested here in the dominant structures. We therefore focus our analysis on the set of conformations with an energy of less than −6.0 kcal/mol. This threshold, which corresponds to an acceptation Metropolis probability of about 0.5% from the lowest-energy structure, serves mostly to remove the highest-energy structures; results presented below are not very sensitive to the exact value of this cutoff. 2809 events in run 1 and 2618 in run 2 satisfy this condition, resulting in a total of 5427 structures. This restriction on low-energy structures includes most structural types. This can be seen, for example, by looking at the average end-to-end distance and radius of gyration: for this subset, these quantities are 9.0 and 5.0 Å, respectively,

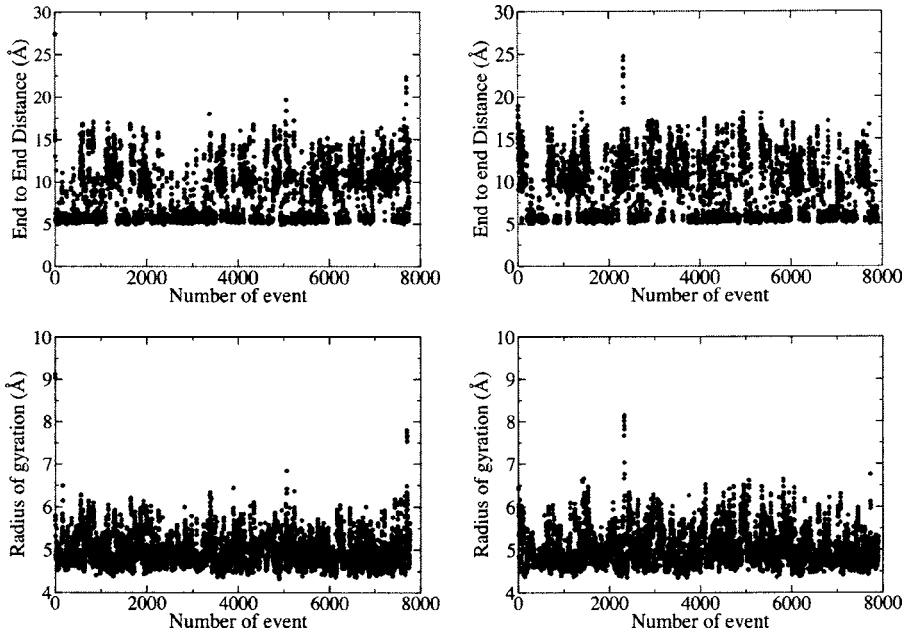


FIG. 2. End-to-end distance in Å (top) and radius of gyration in Å (bottom) as a function of accepted event numbers. Left: data for run 1 and right: data for run 2.

TABLE I. Details of the clusters counting 100 or more structures obtained using a threshold RMSD of 1.0 Å: number (proportion of the total number) of conformations belonging to each cluster, probability of occurrence at $T=298$ K obtained using a Boltzmann weight, and average rms distance measured from the center of the largest cluster.

| Cluster | Number (proportion) of structures | Probability at 298 K | RMSD (Å) |
|---|---|---|---|
| 1 | 1283 (23.6%) | 26.24 | 0.0 |
| 2 | 386 (7.1%) | 1.92 | 1.834 |
| 3 | 321 (5.9%) | 1.64 | 1.807 |
| 4 | 297 (5.5%) | 2.06 | 1.189 |
| 5 | 234 (4.3%) | 1.72 | 4.422 |
| 6 | 233 (4.3%) | 5.02 | 4.728 |
| 7 | 220 (4.1%) | 1.04 | 4.416 |
| 8 | 201 (3.7%) | 9.69 | 4.304 |
| 9 | 122 (2.2%) | 34.35 | 4.454 |
| 10 | 104 (1.9%) | 0.41 | 4.972 |

compared to 9.9 and 4.9 Å, respectively, for all accepted structures.

Clustering analysis on these low-energy structures is done in a recursive manner. The root-mean-square deviation (RMSD) between all structures is calculated and the cluster containing the largest number of conformations within 1 Å is identified. These structures are removed from the list and a new search is performed to identify the second largest cluster. This procedure is repeated until all clusters counting 100 or more structures are found. We have tested the results presented below with a larger threshold of 1.2 Å and found them to be qualitatively the same.

## III. RESULTS

The ART OPEP simulations provide us with a wide set of low-energy structures for $A\beta_{21-30}$. We characterize this conformational ensemble in two steps. First, we look at the lowest-energy structures and identify the dominant clusters found in the simulation. Then we reanalyze the data in terms of the two NMR-derived structural families.[16]

### A. Identification of the lowest-energy structures

The lowest-energy structures in runs 1 and 2 of −9.19 and −9.11 kcal/mol, respectively, are shown in Fig. 1. These two states are referred to as $S^1_{\mathrm{ART}}$ and $S^2_{\mathrm{ART}}$. Structures near these two energy minima are found many times, in both runs, during the simulations: cluster analysis of all structures finds 122 conformations within 1 Å RMSD from $S^1_{\mathrm{ART}}$ and 29 conformations within 1 Å RMSD from $S^2_{\mathrm{ART}}$, indicating that the low-energy structures are well sampled.

We repeat the clustering procedure in order to characterize the other low-energy states and identify the dominant structures in the simulation. All conformations of energy below −6 kcal/mol are clustered using a RMSD threshold of 1.0 Å, as explained in the previous section. The largest cluster includes 23.6% (1283) of all structures. After removing the conformations belonging to this cluster, we redo the clusterization recursively until all clusters of size 100 and above are identified. Note that all these clusters are populated with structures from runs 1 and 2. Since these runs are independent, the two runs of 10 000 events are sufficient to sample all dominant conformational families.

The characterization of the largest clusters can be made using structural and energetic criteria. Some of the details—the cluster size, in terms of the number of structures, and the average rms distance to the center of the largest cluster (cluster 1)—are given in Table I. Table I also gives the Boltzmann-weighted occurrence probability for each cluster $n$,

$$p_n = \frac{\sum_{i \in n} \exp - \beta(E_{i,n} - E_{\min})}{\sum_{\text{all}} \exp - \beta(E_{i,n} - E_{\min})}, \tag{1}$$

where $E_{\min}$ is the lowest-energy structure found in the simulation (−9.19 kcal/mol), $T=298$ K, and $\beta=1/k_BT$. Because the ART-OPEP simulations are not performed in any thermodynamical ensemble, these probabilities are only indicative but they suggest that clusters 1 and 9 are of particular importance; the first one, because of its size (entropy) and the second one, because it contains the lowest-energy $S^1_{\mathrm{ART}}$ structure.

The average distance to the largest cluster (cluster 1) suggests that the ten clusters can be regrouped into a smaller number of structurally similar superclusters. As shown in the last column of Table I, the rms distance to cluster 1 falls roughly into two groups centered at distances of 1.8 and 4.7 Å. Overlapping the representative structure of each cluster, the ten largest clusters are indeed found to belong to three structural families: supercluster 1 (SC1), which includes clusters 1, 2, 3, and 4 [Fig. 3(k)]; supercluster 2 (SC2), with clusters 5, 6, 7, 8, and 10 [Fig. 3(l)]; and the supercluster 3 (SC3) containing cluster 9 alone [Fig. 3(i)]. The probabilities of these superclusters are SC1 (32%), SC2 (17%), and SC3 (34.3%). As seen, all superclusters share a turn between Val24 and Lys28. The SC1 and SC3 superclusters differ, however, in the position of the side chains, while the SC2 supercluster, which contains the $S^2_{\mathrm{ART}}$ minimum, shows a more open structure.

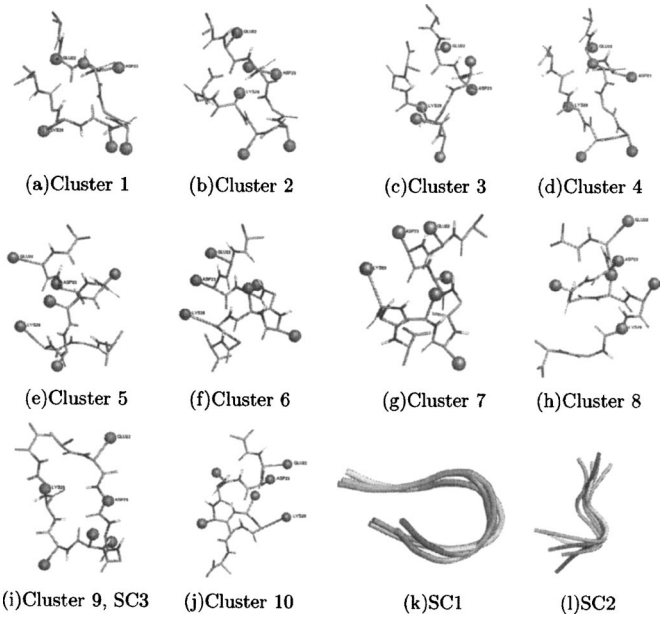


FIG. 3. Cluster and supercluster structures. [(a)–(j)] Centers of the ten clusters counting more than 100 structures and obtained with a threshold of 1.0 Å RMSD, ordered by the number of elements. (k) Supercluster SC1: overlap of clusters 1, 2, 3, and 4; (l) SC2: overlap of clusters 5, 6, 7, 8, and 10; and SC3 is composed of cluster 9 only.

### B. Family analysis

To compare with NMR experiments[16] and previous simulations,[19] we clusterize the conformations around the two NMR-derived families of structures determined, in large part, by the interactions between the three charged polar residues Glu22, Asp23, and Lys28. The main difference between the families lies in the position of the Lys28 side chain, which is placed toward the back of the structure in family I (FI) and toward the front in family II (FII). FI is also characterized by Coulombic interactions between Glu22 and Lys28, while FII is characterized by Coulombic interactions between Asp23 and Lys28.[16] These two conditions can be defined as follows: for family I (family II) (a) an interaction is considered formed between residues 22 (23) and 28 if their side-chain–side-chain distance is $<6.5$ Å and (b) the scalar product between the vector joining $C_\alpha$ and the side chain of Lys28 and the vector joining $C_\alpha$ of Glu22 and $C_\alpha$ of Asp23 is positive (negative), defining the orientation of the Lys28 side chain.

Reanalyzing our datasets using only these geometrical criteria, we find that 1204 ART low-energy structures, 22%, belong to family I and 121, a little more than 2% to family II. While this sorting uses geometrical criteria instead of RMSD, these two families overlap with the two dominant clusters identified in the previous section. More precisely, all structures with FI character belong to cluster 1 [Fig. 3(a)], a member of SC1, and all the structures with FII character, belong to SC3 [Fig. 3(i)], which contains the $S^1_{\mathrm{ART}}$ conformation.

To further characterize the richness of the structures associated with each family,[16,18,19] we follow the evolution of two side-chain–side-chain distances—between Glu22 and Lys28 [denoted by D(2,8)] and between Asp23 and Lys28 [D(3,8)]—and four $C_\alpha \cdots C_\alpha$ distances between amino acids $i$ and $j$ [R$(i,j)$]: R(1,10), R(2,8), R(3,8), and R(4,8). Some of these distances are plotted in Fig. 4 for the ART structures with FI character and in Fig. 5 for the ART structures with FII character. In both cases, we see that only a handful of states dominate, characterized by a narrow range of distances.

In family I, the R(1,10), R(2,8), R(3,8), R(4,8), D(2,8), and D(3,8) distances are centered around 5.6, 5.2, 5.8, 7.9, 6.3, and 9.5 Å, respectively. In particular, while R(2,8) is

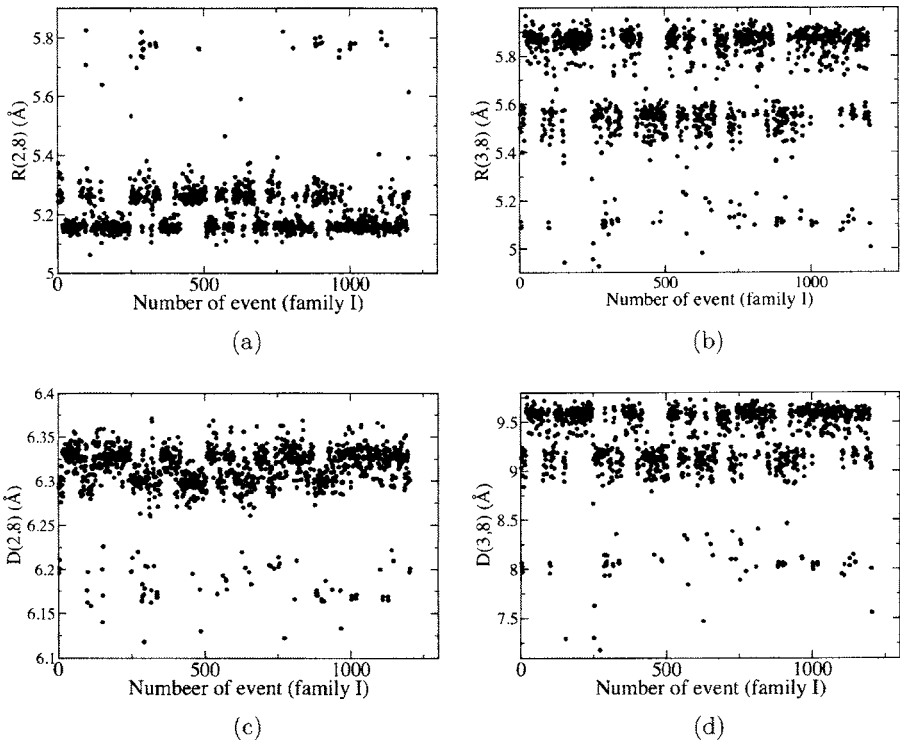


FIG. 4. ART conformations with FI character. Distances between $C_\alpha$ atoms, as described in the text: (a) R(2,8), (b) R(3,8), and between side chains: (c) D(2,8) and (d) D(3,8).

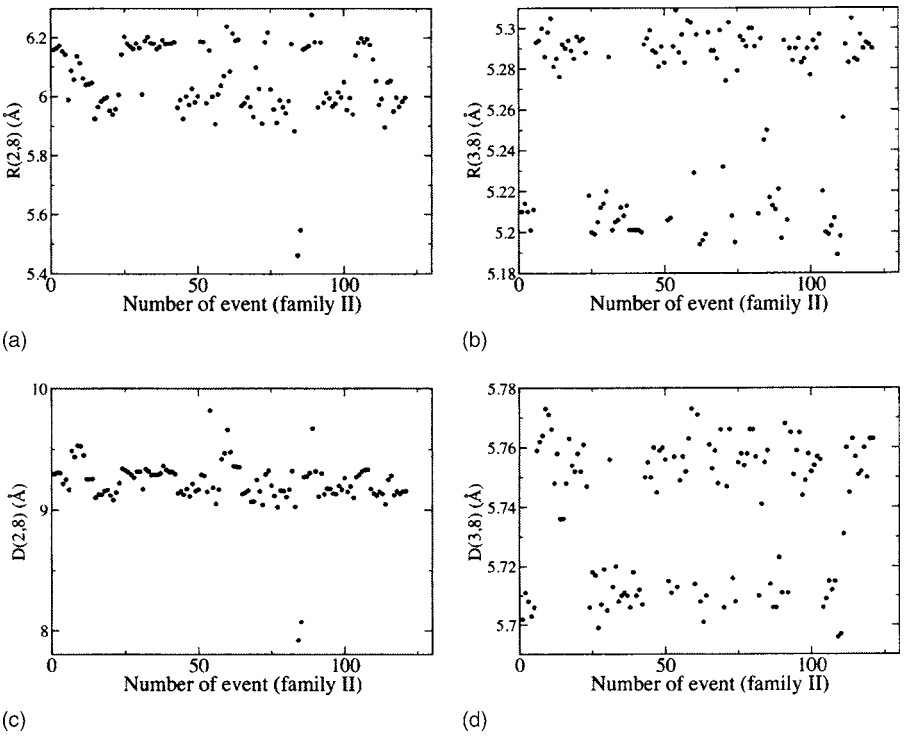


FIG. 5. ART conformations with FII character. Distances between $C_\alpha$ atoms, as described in the text: (a) R(2,8), (b) R(3,8), and between side chains: (c) D(2,8) and (d) D(3,8).

slightly smaller than R(3,8), the difference is much larger between D(2,8) and D(3,8), indicating an attractive salt bridge interaction between the side-chains of Glu22 and Lys28 that stabilizes the turn at Val24-Lys28. In contrast, for family II, the R(1,10), R(2,8), R(3,8), R(4,8), D(2,8), and D(3,8) distances are centered around 5.3, 6.1, 5.3, 6.3, 9.2, and 5.7 Å, respectively. Here, D(3,8) is smaller than D(2,8) and a salt bridge is now formed between Asp23 and Lys28.

Plots in Figs. 4 and 5 also show evidence of some substructures. For example, the R(2,8) distance is stabilized, in ART structures with FI character, around three values: 5.14, 5.29, and 5.80 Å, corresponding to a D(2,8) distance around 6.33, 6.29, and 6.17 Å. As is indicated in Fig. 6, these states correspond to various excitation levels associated with slightly different loop conformations. The structure of the highest-energy structure (shown in light gray) has the tightest loop, with the shortest R(4,8) and D(2,8) distances. As the loop relaxes into place, these distances increase very slightly and the energy decreases by almost 1.5 kcal/mol.

Similarly, we find two substructures in ART structures with FII character, but with a much more subtle position difference, with R(3,8) of 5.21 or 5.29 Å, corresponding to D(3,8) around 5.71 and 5.76 Å, and R(4,8) around 6.15 and 6.27 Å. Here again, these two substructures are associated with a subtle loosening of the loop. As for structures with FI character, small variations lead to an energy increase of 1.5 kcal/mol or even more (Fig. 7).

## IV. DISCUSSION

### A. Comparison with previous simulations

$A\beta_{21-30}$ has been the subject of simulations using REMD with explicit solvent,[19] standard MD with explicit solvent,[18] and DMD with explicit solvent.[20] All simulations report, similarly to what we find here, a loop spanning Val24-Lys28 with the possibility for the side chain of Lys28 to be oriented above or below the plane of the structure. In addition to this general agreement, Baumketner *et al.* also identify two REMD-generated dominant clusters and offer a direct comparison with the experimentally measured ROEs.[19] The dominant cluster $C1^{REMD}$, 30% populated at 300 K, fits the geometrical constraints of family I as defined by NMR, but the cluster $C2^{REMD}$, 10% populated, does not meet the constraints of family I or II.

By calculating the $C_\alpha$ RMSd between all ART-generated structures of energy less than −6 kcal/mol and the two dominant REMD-generated clusters C1 and C2,[19] using residues 22–28 because the extremities are rather flexible, we find

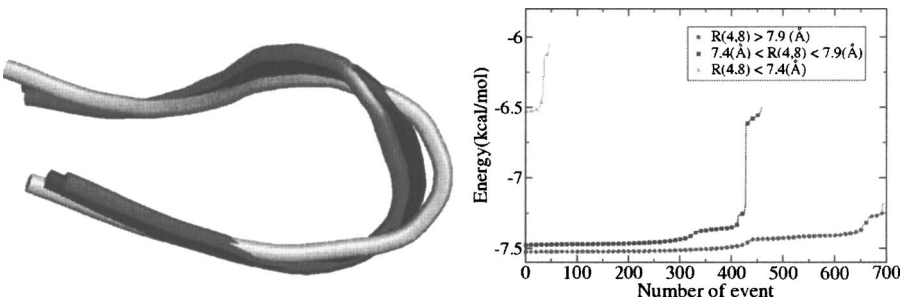


FIG. 6. Left: conformations corresponding to the three substructures identified from Fig. 4, for FI family. Going from light gray to black and dark gray, the size of the loop [characterized by R(4,8)] increases slightly, allowing a better relaxation of the overall structure. Right: the energy in increasing order for the FI family structures and classified by the R(4,8) distance. The color of the points match that of the structures on the left.

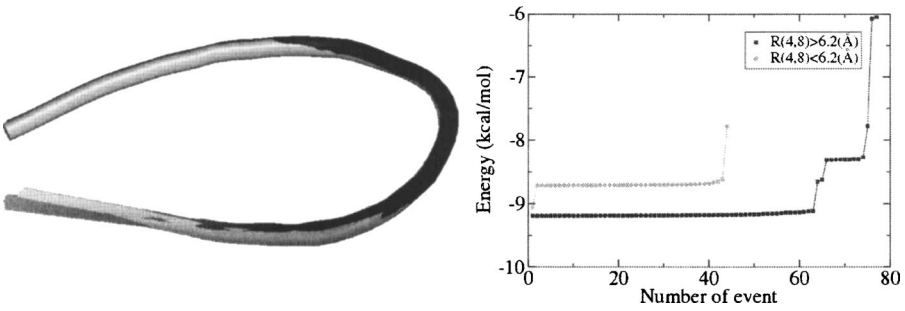


FIG. 7. Left: conformations corresponding to the two substructures identified from Fig. 5, for FII family. Going from gray to black, the size of the loop [characterized by R(4,8)] increases slightly, allowing a better relaxation of the overall structure. Right: the energy in increasing order for the FII family structures and classified by the R(4,8) distance. The color of the points match that of the structures on the left.

ART conformations deviating by 0.3 Å RMSD from $C1^{REMD}$ and by 2.6 Å RMSD from $C2^{REMD}$. By limiting the RMSD comparison to the center of the ten clusters, we find that $C1^{REMD}$ is at 2.1 Å from cluster 5 and at 2.2 Å from cluster 9. It can therefore be classified as being halfway between the superclusters SC2 and SC3. Similarly, $C2^{REMD}$ is at 2.9 Å from cluster 8 and 3.1 Å from cluster 5, but at 3.6 Å from cluster 1 and can be defined as being between superclusters SC1 and SC2.

Taken together, these results indicate that there is a good structural overlap between the dominant superclusters $C1^{REMD}$ and SC3 generated by REMD and ART-OPEP simulations, respectively, but the other clusters differ substantially.

### B. Comparison with NMR experiments

We have already established that the two most likely clusters, clusters 1 and 9, satisfy the NMR-derived geometrical criteria associated with families I and II. We now verify that the ART OPEP structures are near the two NMR-derived models associated with these families (NMR-1 and NMR-2).[16] By using the $C_\alpha$ atoms of residues 21–30, we find that the smallest RMSD between the NMR-1 (NMR-2) structure and all ART conformations below −6 kcal/mol is 1.6 Å (1.9 Å). Restricting the set to the centers of the ten largest clusters, NMR-2 is at 2.5 Å from cluster 9 (1.9 Å using residues 22–28), and NMR-1 is at 3.4 Å from cluster 1 (2.2 Å using residues 22–28).

We can also reverse the analysis and compute from our simulations the interproton distances, identifying directly the contacts measured by NMR at 10 °C. Four medium range $(i,i+2)$ constraints and one long-range constraint involving residues 22 and 30 are found by NMR.[16] To this end, H$\alpha$ atoms are added to the ART-generated structures using standard geometrical parameters. Following Ref. 19 the interproton distances are then averaged using a $\langle -1/r^6\rangle^{-1/6}$ weight to reproduce the NMR signal. Because the strength of the ROEs decreases rapidly, distances longer than about 5.0 Å are not visible by NMR.

Table II shows the calculated interproton distances in our models, the dominant $C1^{REMD}$ cluster and the NMR-1 and NMR-2 structures. We see that both SC1, associated with NMR-1, and SC3, associated with NMR-2, respect the five constraints (although the 22$\alpha$–24N constraint is just at the limit for SC3). This is in contrast to the $C1^{REMD}$ cluster which violates two distances by 1.5 Å. The situation is different for SC2: as it is stretched out, it is not surprising that it violates the constraint associated with a closed loop (22$\alpha$–30N), interestingly, however, it does respect the four local constraints. Overall, the agreement with NMR suggests that our simulations provide a reliable picture of the conformational space of A$\beta_{21-30}$ in solution.

Several studies have emphasized the role of the intramolecular salt bridge between Asp23 and Lys28 in fibril formation.[21,36] A recent solid-state NMR-derived model of A$\beta_{1-40}$ also indicates a turn conformation spanning residues 21–30.[6] By using the $C_\alpha$ atoms of residues 22–28, we find that the NMR-1, NMR-2, and lowest-RMSD ART OPEP structures deviate by 2.3, 2.3, and 1.2 Å from the conformation within the fibril model. However, the corresponding ART OPEP conformation is destabilized by 5.0 kcal/mol with respect to the lowest-energy $S^1_{ART}$ structure. This emphasizes the role of adjacent residues within A$\beta$ or the impact of polymerisation in stabilizing the region 21–30 within a fibril-competent state.

## V. CONCLUSIONS

We have determined the structural properties of the fragment A$\beta_{21-30}$ from two independent ART OPEP simulations. By clustering and visualizing the low-energy structures, we

TABLE II. Weighted-averaged interproton distances using our three superclusters, the two NMR-derived structures (Ref. 16) and the dominant cluster C1 generated by REMD (Ref. 19). In bold, distances larger than 5 Å, violating the constraints.

| ROEs | 21$\alpha$–23N | 22$\alpha$–24N | 24$\alpha$–26N | 28$\alpha$–30N | 22$\alpha$–30N |
|---|---|---|---|---|---|
| SC1-ART | 3.12 | 4.04 | 4.46 | 4.42 | 4.77 |
| SC2-ART | 3.31 | 3.59 | 4.64 | 4.94 | **7.42** |
| SC3-ART | 4.12 | 5.03 | 4.69 | 4.69 | 4.22 |
| NMR-1 | 4.79 | 3.64 | 4.63 | 4.55 | **5.38** |
| NMR-2 | 4.92 | 3.54 | 4.33 | **5.07** | 4.84 |
| C1-REMD | 4.97 | **6.36** | **5.38** | 4.36 | **6.53** |

find that $A\beta_{21-30}$ samples a narrow set of ten structural basins which can be further clustered into three superclusters, all involving a loop from residues Val24 to Lys28. The superclusters SC1 and SC3 differ mostly in the orientation of the side groups; the supercluster SC2 shows a more open structure.

The dominant clusters, as calculated using a simple Boltzmann energy reweighting, overlap perfectly with the two families of structures, FI and FII, obtained by enforcing a set of geometrical constraints extracted from NMR experiments.[16] All structures with FI geometrical character are part of the dominant cluster, as defined by the number of its members, itself belonging to the supercluster SC1. Similarly, all conformations with FII character belong to the supercluster SC3 with the lowest-energy conformation. These two superclusters also respect the five interproton constraints found by ROESY NMR spectrum,[16] providing a significantly better agreement than REMD simulations. In contrast, the supercluster SC2, which contains the second lowest-energy structure, meets the four local NMR constraints, but has a distance between the backbone atoms of Glu22 and Ala30 which cannot be detected by ROESY experiments. Given the agreement between the (SC1, SC3) superclusters with NMR, we believe that SC2 is also a true local free energy minimum.

Our simulations indicate the presence of a turn spanning the residues Val24-Lys28 in all dominant structures of $A\beta_{21-30}$ in solution. This result, fully consistent with NMR experiments and, to various degrees, with previous simulations, further supports the hypothesis that the persistence of this turn may explain the resistance of the region (21–30) to proteolysis in the $A\beta_{40}$ and $A\beta_{42}$ peptides.[16] Finally, this study, by reproducing $A\beta_{21-30}$ NMR data, provides a strong basis for an extensive ART OPEP study of the structural properties of full-length $A\beta$ peptides.

## ACKNOWLEDGMENTS

The authors thank Dr. David Teplow for providing us with the pdb files of the two NMR-derived families and REMD-generated clusters. The authors (W.C., P.D., and N.M.) are supported in part by the Alzheimer Society of Canada. One of the authors (N.M.) also acknowledges partial support from Natural Sciences and Engineering Research Council of Canada, the Canada Research Chair Fund, and the Fonds Québécois de Recherche sur la Nature et les Technologies. Another author (P.D.) acknowledges fundings from CNRS and Université of Paris 7. Most of the calculations were done on the computers of the Réseau Québécois de Calcul de Haute Performance (RQCHP).

[1] R. Wetzel, Adv. Protein Chem. **50**, 183 (1997).
[2] A. L. Fink, Folding Des. **3**, R9 (1998).
[3] D. Thirumalai, D. Klimov, and R. Dima, Curr. Opin. Struct. Biol. **13**, 146 (2003).
[4] G. Bitan, M. D. Kirkitadze, A. Lomakin, S. S. Vollers, G. B. Benedek, and D. B. Teplow, Proc. Natl. Acad. Sci. U.S.A. **100**, 330 (2003).
[5] G. Bitan, S. S. Vollers, and D. B. Teplow, J. Biol. Chem. **278**, 34882 (2003).
[6] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko, Proc. Natl. Acad. Sci. U.S.A. **99**, 16742 (2002).
[7] T. Lührs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert, and R. Riek, Proc. Natl. Acad. Sci. U.S.A. **102**, 17342 (2005).
[8] J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed, and R. Tycko, Biochemistry **45**, 13748 (2000).
[9] B. Y. Ma and R. Nussinov, Proc. Natl. Acad. Sci. U.S.A. **22**, 14126 (2002).
[10] G. Boucher, N. Mousseau, and P. Derreumaux, Proteins: Struct., Funct., Bioinf. (to be published).
[11] J. Kim and M. Lee, Biochem. Biophys. Res. Commun. **316**, 393 (2004).
[12] W. Han and Y.-D. Wu, J. Am. Chem. Soc. **127**, 15408 (2005).
[13] T. L. S. Benziger, D. M. Gregory, T. S. Burkoth, H. Miller-Auer, D. G. Lynn, R. E. Botto, and S. C. Meredith, Proc. Natl. Acad. Sci. U.S.A. **95**, 13407 (1998).
[14] J.-T. Guo, R. Wetzel, and Y. Xu, Proteins: Struct., Funct., Bioinf. **57**, 357 (2004).
[15] J. Talafous, K. J. Marcinowski, G. Klopman, and M. G. Zagorski, Biochemistry **33**, 7788 (1994).
[16] N. D. Lazo, M. A. Grant, M. C. Condron, A. C. Rigby, and D. B. Teplow, Protein Sci. **14**, 1581 (2005).
[17] L. Hou, H. Shao, Y. Zhang *et al.*, J. Am. Chem. Soc. **126**, 1992 (2004).
[18] L. Cruz, B. Urbanc, J. M. Borreguero, N. D. Lazo, D. B. Teplow, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **102**, 18258 (2005).
[19] A. Baumketner, S. L. Bernsein, T. Wyttenbach, N. D. Lazo, D. B. Teplow, M. T. Bowers, and J.-E. Shea, Protein Sci. **6**, 1239 (2006).
[20] J. M. Borreguero, B. Urbanc, S. V. Buldyrev, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **102**, 6015 (2005).
[21] K. L. Sciarretta, D. J. Gordon, A. T. Petkova, R. Tycko, and S. C. Meredith, Biochemistry **44**, 6003 (2005).
[22] G. T. Barkema and N. Mousseau, Phys. Rev. Lett. **77**, 4358 (1996).
[23] R. Malek and N. Mousseau, Phys. Rev. E **62**, 7723 (2000).
[24] N. Mousseau, P. Derreumaux, G. T. Barkema, and R. Malek, J. Mol. Graphics Modell. **19**, 78 (2001).
[25] J. Maupetit, P. Tuffery, and P. Derreumaux (unpublished).
[26] S. Santini, G. H. Wei, N. Mousseau, and P. Derreumaux, Structure (London) **12**, 1245 (2004).
[27] S. Santini, N. Mousseau, and P. Derreumaux, J. Am. Chem. Soc. **126**, 11509 (2004).
[28] A. Melquiond, N. Mousseau, G. Boucher, and P. Derreumaux, J. Chem. Phys. **122**, 174904 (2005).
[29] G. H. Wei, N. Mousseau, and P. Derreumaux, Biophys. J. **87**, 3648 (2004).
[30] P. Derreumaux, J. Chem. Phys. **111**, 2301 (1999).
[31] G. H. Wei, P. Derreumaux, and N. Mousseau, J. Chem. Phys. **119**, 10712 (2003).
[32] G. H. Wei, N. Mousseau, and P. Derreumaux, J. Phys.: Condens. Matter **16**, S5047 (2004).
[33] N. S. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[34] P. Derreumaux, Phys. Rev. Lett. **85**, 206 (2000).
[35] G. H. Wei, N. Mousseau, and P. Derreumaux, J. Chem. Phys. **117**, 11379 (2002).
[36] A. T. Petkova, W. M. Yau, and R. Tycko, Biochemistry **45**, 489 (2006).