Amyloid oligomer structure characterization from simulations: A general method

Phuong H. Nguyen,[1,a)] Mai Suan Li,[2] and Philippe Derreumaux[1,3,b)]

[1]*Laboratoire de Biochimie Théorique, UPR 9080, CNRS Université Denis Diderot, Sorbonne Paris Cité IBPC, 13 rue Pierre et Marie Curie, 75005 Paris, France*
[2]*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland*
[3]*Institut Universitaire de France, 103 Bvd Saint-Germain, 75005 Paris, France*

Amyloid oligomers and plaques are composed of multiple chemically identical proteins. Therefore, one of the first fundamental problems in the characterization of structures from simulations is the treatment of the degeneracy, i.e., the permutation of the molecules. Second, the intramolecular and intermolecular degrees of freedom of the various molecules must be taken into account. Currently, the well-known dihedral principal component analysis method only considers the intramolecular degrees of freedom, and other methods employing collective variables can only describe intermolecular degrees of freedom at the global level. With this in mind, we propose a general method that identifies all the structures accurately. The basis idea is that the intramolecular and intermolecular states are described in terms of combinations of single-molecule and double-molecule states, respectively, and the overall structures of oligomers are the product basis of the intramolecular and intermolecular states. This way, the degeneracy is automatically avoided. The method is illustrated on the conformational ensemble of the tetramer of the Alzheimer's peptide $A\beta_{9-40}$, resulting from two atomistic molecular dynamics simulations in explicit solvent, each of 200 ns, starting from two distinct structures. © 2014 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4866902]

I. INTRODUCTION

One hallmark of Alzheimer's disease is the formation of amyloid plaques in the brain, resulting from the self-assembly of $A\beta$ peptides of various amino acid lengths, though $A\beta_{1-40}$ is the dominant species and $A\beta_{1-42}$ is the most toxic.[1]

Structural and dynamical characterization of the oligomers is challenging by experimental means due to their metastability and high aggregation propensity. As a result, these oligomers are not amenable to solution NMR and X-ray crystallography and only low-resolution data are available.[2,3]

Understanding the physical factors driving amyloid formation is also a challenge by computer simulations due to the intrinsic disorder of the peptides and the size of the energy landscape to be explored. All-atom, coarse-grained, and mesoscopic representations coupled to state-of-the-art methods can provide insights into the transition from amorphous aggregates to fibrils.[4–23]

The data analysis of the aggregates is also a difficult task. The radius of gyration,[5] the β-sheet size,[12,13] the oligomer size distribution,[14,15,17] the number of contacts,[13,14] the order parameter P_2,[21] the angles between β-sheets,[22] the connectivity length,[12] and the chain-independent and/or orientation independent C_α root-mean-square deviation and cluster analysis[6,7,12] are often used to characterize the structures, but each variable captures only one feature of the self-assembly. Even with the use of more complex collective variables,[24] the characterization and classification of the conformations remain difficult for highly flexible peptides, large and heterogeneous aggregates and systems characterized by degenerate states, that is, identical oligomer conformations differing only by the permutations of the molecules.

The dihedral-angle principal component analysis (dPCA), originally developed for single proteins,[25] is often used to construct the free energy landscapes (FEL) of oligomers.[11,21,26–30] Because the dPCA uses the dihedral angles of monomers, it only captures the intramolecular degrees of freedom of the oligomers. As a first step toward including the intermolecular degrees of freedom, an angle for each pair of monomers that accounts for their relative orientation was included in dPCA.[31] This additional information did not improve, however, the resolution of the FEL of $A\beta_{16-22}$ trimer because the dPCA does not treat correctly the degenerate states.[6,7,12,14,15,32]

To understand the effect of degeneracy on the FEL, let us consider a toy system consisting of two identical particles whose positions are generated by two independent trajectories: $X_1(t_n) = r_1 + 1$, $X_2(t_n) = r_2 + 4$ for the first trajectory, and $X_1(t_n) = r_3 + 1$, $X_2(t_n) = r_4 + 4$ for the second trajectory, where r_i is a random number $\in (0, 1)$ [Fig. 1]. If the particles are treated as distinguishable objects in the analysis, and their labels X_1, X_2 are fixed in time and in the two trajectories [Figs. 1(a) and 1(b)], then the time-averaged position of each particle is easily obtained: $\langle X_1 \rangle = 1.5$, $\langle X_2 \rangle = 4.5$. However, because particles are indistinguishable, one can perform a permutation of the label of the particles, for example, in the trajectory 2 as: $X_1 \rightarrow X_2$, $X_2 \rightarrow X_1$ [Figs. 1(d)

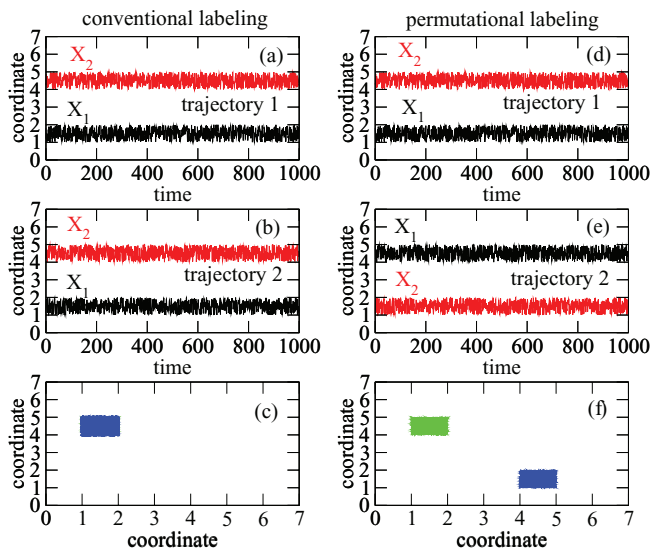a)Email: phuong.nguyen@ibpc.fr
b)Email: philippe.derreumaux@ibpc.fr

FIG. 1. Trajectories of a toy system consisting of two identical particles with fixed labels X_1, X_2 (a) and (b) and with permutated labels (d) and (e). The free energy landscape $F(X_1, X_2)$ obtained from the two trajectories where the label of two particles is fixed (c) and permutated (d).

and 1(e)], without changing the structure of the system. However, the time-averaged position of each particle is affected by this label permutation: $\langle X_1 \rangle = 3.0$ and $\langle X_2 \rangle = 3.0$. Next, we construct the FEL of the system $F(X_1, X_2) = -k_B T \ln P(X_1, X_2)$, where $P(X_1, X_2)$ is the joint probability distribution obtained from a histogram of the data of the two trajectories. If the particles are treated as distinguishable objects and their labels are fixed in time and in two trajectories, then the FEL exhibits only one state, which is the superposition of two trajectories [Fig. 1(c)], whereas the FEL splits into two states with the permutation of the particle labels [Fig. 1(f)]. From this toy example, one can see the impact of degenerated states on FEL.

Recently, a method that considers both the intramolecular and intermolecular degrees of freedom and correctly treats the degeneracy was proposed by Riccardi *et al.*,[32] and applied to construct the FEL of the $A\beta_{16-22}$ trimer. Basically, the overall state of the trimer is defined as the product basis of the intermolecular and intramolecular states. The intramolecular states of the trimer are described in terms of the structure of single monomers obtained by the dPCA, thus avoiding the degeneracy problem. To account for the relative orientation of any two peptides and because $A\beta_{16-22}$ trimer tends to form full extended conformations, Riccardi *et al.*[32] used the angle θ between end-to-end vectors and associated the values $0° \leq \theta \leq 50°$ with a parallel orientation, $(130° \leq \theta \leq 180°)$ with an antiparallel orientation, and other values with random. Using this information for all pairs of peptides, the degenerate states of the trimer were treated properly.[32]

There are, however, two limitations with the use of the angle θ. First, this angle cannot describe the intermolecular structure of oligomers having arbitrary shapes and arrangements of the peptides. Second, we do not know in advance the structures of oligomers. Here, we propose an automatic method that identifies all the structures accurately while treating the degenerate states, independently of the shapes of the

oligomers. The intramolecular and intermolecular states are described in terms of combinations of single-molecule and double-molecule states, respectively, and following Ref. 32, the overall structures of oligomers are the product basis of the intramolecular and intermolecular states. The method is illustrated on the conformational ensemble of the tetramer of the Alzheimer's peptide $A\beta_{9-40}$, resulting from two molecular dynamics (MD) simulations in explicit solvent, each of 200 ns, starting from two distinct structures. This biological system of interest was chosen because it has many intermolecular and intramolecular degrees of freedom. While our simulation goes on step beyond previous MD simulations of $A\beta_{1-42}$ and $A\beta_{9-40}$ trimers and pentamers starting from fully formed fibrils for 100 ns,[33,34] we emphasize that our two independent trajectories, each of 200 ns, do not converge to equilibrium within the current simulation time. They allow us, however, to explain the essential aspects of the method and the results in details.

II. THEORY

A. Simulation details

We used the GROMOS96 force field 43a1[35] to model the tetramer system and the SPC (Simple Point Charge) water model[36] to describe the solvent. The $A\beta_{9-40}$ structure was taken from the solid-state NMR structure of $A\beta_{1-40}$ fibrils with two β-strands at positions 10–23 and 30–38.[37] Two independent MD simulations were performed. In the first one, the four peptides are distant but have a parallel orientation. In the second simulation, the four monomers have random orientations and distances between their centers of mass [Fig. 2]. The systems were solvated in cubic boxes containing about 30 000 water molecules.

The GROMACS program suite[38,39] was employed. The equations of motion were integrated by using a leap-frog
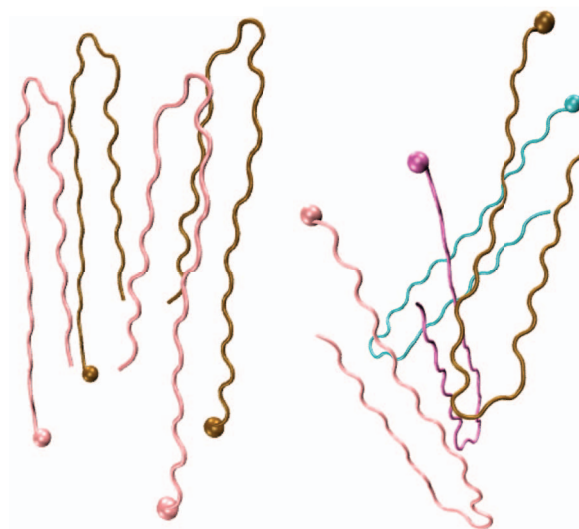

FIG. 2. Initial structures of $(A\beta_{9-40})_4$ used in the MD simulations. (Left) Monomers belonging to the same layer have distinct colors and are separated by 1.4 nm. The distance between the two layers is 2.5 nm. (Right) Random orientations and positions of the peptides. The shortest and longest distances between the centers of mass of two monomers are 1.4 and 3.2 nm, respectively. The ball indicates the first residue.

algorithm with a time step of 2 fs. Covalent bond lengths were constrained via the SHAKE[40] procedure with a relative geometric tolerance of 10^{-4}. We used the particle-mesh Ewald method to treat the long-range electrostatic interactions[41] and updated the non-bonded interaction pair-list every 5 fs.

The systems were minimized using the steepest decent method. Subsequently, the solvated system was equilibrated for 100 ps at constant pressure (1 atm) and temperature ($T = 300$ K), respectively, using the Berendsen coupling procedure.[42] The system was then equilibrated further at constant temperature ($T = 300$ K) and constant volume for 1 ns. Each trajectory was then run for 200 ns and configurations were saved every 2 ps.

B. Data analysis

1. Secondary structure, contacts, β-sheet, and k-means clustering

We monitored the secondary structures of the tetramer using the STRIDE program.[43] We grouped the eight secondary structures defined by STRIDE (3_{10}-helix, PI-helix, extended, bridge, turn, coiled, and bend) into four main structures: beta = extended + bridge, alpha = helix + 3_{10}-helix + PI-helix, turn = turn + bend, and coil = coil. A contact between two side-chains was formed if the distance between their centers of mass is less than 0.65 nm. We also described the oligomers in terms of p-stranded $β$-sheets, with p the number of $β$-strands consecutively connected by at least two backbone hydrogen bonds (H-bonds). A $β$-strand was defined if at least any three consecutive residues are in the extended conformation, and a H-bond was formed if the distance between the N and O atoms is less than 0.35 nm and the angle (CON) is greater than $120°$.

To identify the conformational states on the FEL, we carried out clustering analysis using the Hartigan-Wong k-means algorithm[44] as implemented in the R program suite.[45] As the k-means clustering is sensitive to the initial conditions, the algorithm was run 2000 times, and the best results were chosen. To determine the optimal value of the number of clusters, we examined the sum of squares of the observations to their assigned cluster centers and the 30 most popular indices for cluster validation provided by the NbClust package[46] and we performed a visual inspection of the partitioning. All structures shown correspond to the centers of clusters.

2. Identification of the intermolecular states

Assuming that the oligomer is simulated for N_s steps, and is composed of N_c chemically identical chains, each chain consisting of N_r residues, there are $i = 1, \ldots, N_c \times (N_c - 1)/2$ double-chains (or pairs of chains) and each double-chain contains $m = 1, \ldots, N_r \times N_r$ distances between the inter-chain centers of mass of the side-chains.

We characterize the *intermolecular states* of the oligomer in terms of double-molecule states which describe the position-orientation arrangements of any two chains. To this end, let $[d_i^m(t_n)]$ be the distance of the mth inter-chain side-chain pair of the ith double-chain measured at time t_n ($n = 1, \ldots, N_s$). Now, we map all these distances to

$[d_i^m(t_n)] \mapsto [D_m(t_j)]$, $m = 1, \ldots, N_r \times N_r$, $j = (n - 1) \times N_c \times (N_c - 1)/2 + i$. This new representation can be imagined as the distances between the inter-chain centers of mass of the side-chains for two peptides simulated alone during $N_c \times (N_c - 1) \times N_s/2$ steps.

A knowledge of the set $[D_m(t_j)]$ allows us to fully describe all positions and orientations of all the chains in the oligomer. That is, we can obtain the intermolecular states of the oligomer in terms of combinations of the structures of double-chains, thereby avoiding the degeneracy. These structures can be obtained through the PCA of the inverse distances $1/D_m(t_j)$.[47] The PCA is carried out by calculating the covariance matrix

$$\sigma_{ab} = \langle (D^a - \langle D^a \rangle)(D^b - \langle D^b \rangle) \rangle, \qquad (1)$$

with $a, b = 1, \ldots, N_r \times N_r$ and $\langle \ldots \rangle$ denotes the average over all sampled double-chains of a virtual trajectory of length $N_c \times (N_c - 1) \times N_s/2$. By diagonalizing σ, we obtain $N_r \times N_r$ eigenvectors $\mathbf{v}^{(l)} = \{v_{nl}\}$ (v_{nl} is the nth component of the lth eigenvector) and eigenvalues λ_n, which are rank-ordered in descending order, i.e., λ_1 represents the largest eigenvalue. The eigenvectors and eigenvalues of σ yield the modes of collective motion and their amplitudes. The lth principal component is defined as $V_l = \mathbf{v}^{(l)} \cdot \mathbf{q}$. It has been shown that a large part of the system's fluctuations can be described in terms of only a first few principal components.[48–51] The FEL spanned by the first p principal components $V = (V_1, \ldots, V_p)$ is given by $F_{\text{double}}(V) = -k_B T[\ln P(V) - \ln P_{\text{max}}]$, where $P(V)$ is the probability distribution obtained from a histogram of the MD data.

Let us call the states s_d^k ($k = 1, \ldots, n_{\text{state}}$) the *double-molecule* states on $F_{\text{double}}(V_1, \ldots, V_p)$. The positions and orientations of the double-molecules are known from the trajectory, thus, an intermolecular state, S_{inter}^l (l denotes the state's index), of the oligomer is completely characterized by $N_c \times (N_c - 1)/2$ double-molecule states s_d^k. In principle, there are $(n_{\text{state}})^{N_c \times (N_c-1)/2}$ possible combinations of s_d^k states that form states S_{inter}^l. In practice, there are degenerated states. For example, let us consider a case where $N_c = 3$ chains, i.e., there are $3 \times 2/2 = 3$ double-chains, and each pair samples two double-molecule states ($n_{\text{state}} = 2$): s_d^1 and s_d^2. In total there are 8 possible combinations of s_d^k. However, the combinations (s_d^1, s_d^1, s_d^2), (s_d^1, s_d^2, s_d^1), and (s_d^2, s_d^1, s_d^1) represent the same oligomer structure because the chains are indistinguishable, and after removing the degeneracy, there are 4 unique combination states.

3. Identification of the intramolecular states

We characterize the *intramolecular states* of the oligomers in terms of the structures of single monomers. This idea was first proposed in our recent work.[32] To this end, let $[\phi_i^m(t_n), \psi_i^m(t_n)]$ ($m = 1 \ldots N_r$, $i = 1 \ldots N_c$) be the backbone dihedral angles of the residue mth of the chain ith. Again, we map these angles to new variables $[\phi_i^m(t_n), \psi_i^m(t_n)] \mapsto [\Phi^m(t_j), \Psi^m(t_j), j = (n-1)N_c(N_c - 1)/2 + i$. These variables can be imagined as the angles of a monomer simulated alone for $N_c \times N_s$ steps. A knowledge of the set of the angles $[\Phi^m(t_j), \Psi^m(t_j)]$ allows us to fully describe all possible

intramolecular states of all chains in the oligomer. That is, we can obtain intramolecular states of the oligomer in terms of the structures of single-chains, thereby avoiding the degeneracy problem. The structures of single chains can be obtained through the dPCA.[25] Briefly, we define the covariance matrix

$$\sigma_{ab} = \langle (q_a - \langle q_a \rangle)(q_b - \langle q_b \rangle) \rangle, \tag{2}$$

with $q_c = \cos(\alpha_c)$ and $q_{c+1} = \sin(\alpha_c)$. Here, $\alpha_c \in \{\Phi^m, \Psi^m\}$, $m = 1 \ldots 2 \times N_r$ and $\langle \ldots \rangle$ denotes the average over all sampled conformations of the single-chain trajectory of length $N_c \times N_s$. The diagonalization of the covariance matrix provides the principal components which are then used to construct the FEL, F_{single}, of a single monomer. The FEL spanned by the first p principal components $V = (V_1, \ldots, V_p)$ is given by $F_{single}(V) = -k_B T[\ln P(V) - \ln P_{max}]$, where $P(V)$ is the probability distribution obtained from a histogram of the MD data.

Let us call s_s^k ($k = 1, \ldots, n_{state}$) the *single-molecule* states on $F_{single}(V_1, \ldots, V_p)$. The structures of single chains are known from the trajectory, thus, an intramolecular state, S_{intra}^m (m denotes state's index), of the oligomer is completely characterized by N_c single-molecule states s_s^k. In principle, there are $(n_{state})^{N_c}$ possible combinations of s_s^k states that form states S_{intra}^m. In practice, there is degeneracy in those states, and not all these states are sampled during the simulation.

4. Identification of the overall oligomer states

Having constructed the intermolecular states S_{inter}^l and the intramolecular states S_{intra}^m, the overall states of the oligomer can be obtained from the product basis $S_{overall}^n = S_{inter}^l \times S_{intra}^m$ (n denotes the state's index).

III. RESULTS AND DISCUSSION

We have discussed the general effect of degeneracy on the FEL of a toy system. We now show how to obtain the structures of the tetramer formed by the $A\beta_{9-40}$ peptide by taking into account correctly (i) the degenerate states and (ii) both the intermolecular and intramolecular degrees of freedom.

A. Intramolecular states of $(A\beta_{9-40})_4$

Each configuration of the tetramer consists of 128 pairs of the backbone dihedral angles, which specify completely the intramolecular structures of the tetramer. Now, we map these 128 pairs to 32 pairs: $[\phi_i^m(t_n), \psi_i^m(t_n)] \mapsto [\Phi^m(t_j), \Psi^m(t_j)]$, ($m = 1 \ldots 32$, $i = 1 \ldots 4$), which are considered as angles sampled by a virtual single-chain simulated alone at times $t_j = t_n$, $t_n + 1$, $t_n + 2$, and $t_n + 3$. Now, the dPCA can be directly applied to these new angles $[\Phi^m, \Psi^m]$ without the degeneracy problem. The FEL projected on the first two principal components, which account for about 65% of the system's fluctuation, is shown in Fig. 3. Analysis using k-means clustering method[44] reveals 5 single-molecule states. Again, note that these states are expected to be different from those sampled by the peptide simulated alone, because the interpeptide interactions shift the equilibrium structures.[52] The conforma-
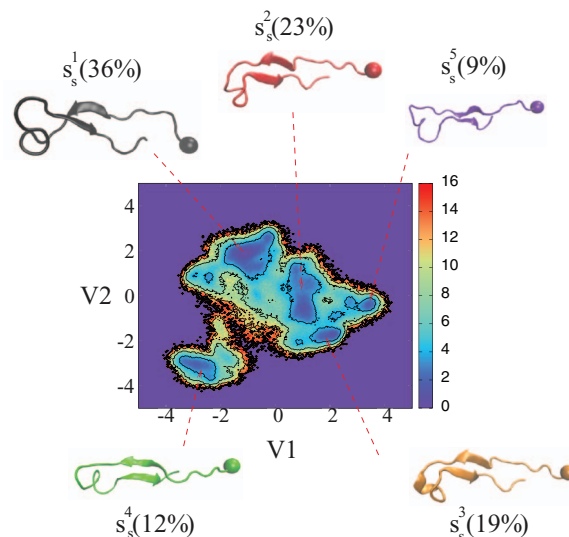

FIG. 3. Free energy landscape (in kcal/mol) of a single $A\beta_{9-40}$ molecule obtained from MD simulations of $(A\beta_{9-40})_4$ projected on the first two principal components V1 and V2 obtained from dPCA. The center of each metastable single-molecule state is shown in a different color. The ball indicates the first residue.

tion of each $A\beta_{9-40}$ chain can switch between these 5 states. Table I characterizes the molecular structures of these five states.

First, the α-helix is marginally populated ($\leq 1\%$) in all states. This is not surprising because the simulations started from isolated monomers with fibril-like structure and were run for 200 ns, and the GROMOS force field is known to favor β-strand structure.[30] Overall, we see that the five states are characterized by deformed β-hairpins with a β-strand content varying between 23% and 40%, a coil content between 30% and 42% and a turn content at 28%–47%. The maximal (minimal) lengths of the β-strands cover residues 11–15 (11–12), 16–23 (16–19) and 30–37 (35–36) in s_s^1, residues 14–22 (14–15) and 29–36 (34–36) in s_s^2, residues 11–14 (11–12), 16–20 (16–17), 24–26 and 30–39 (32–33) in s_s^3, residues 18–21 and 33–35 in s_s^4, and residues 16–19, 34–37 in s_s^5. It is interesting to note that all MD-generated β-strands span the hydrophobic core (CHC, 17–21 residues) and the residues 32–35, while the β-strands span the residues 10–23 and 30–38 in the initial structures.

TABLE I. (Left) Structural characterization of the single-molecule states and the intramolecular states of the tetramer. For the single-molecule states, s_s^k ($k = 1, \ldots, 5$), shown in Fig. 3 we give their populations P (in %) and secondary structure contents (in %). (Right) For each intramolecular state, S_{intra}^m ($m = 1, \ldots, 5$), of $(A\beta_{9-40})_4$, we give its populations and its combination of single-molecule states.

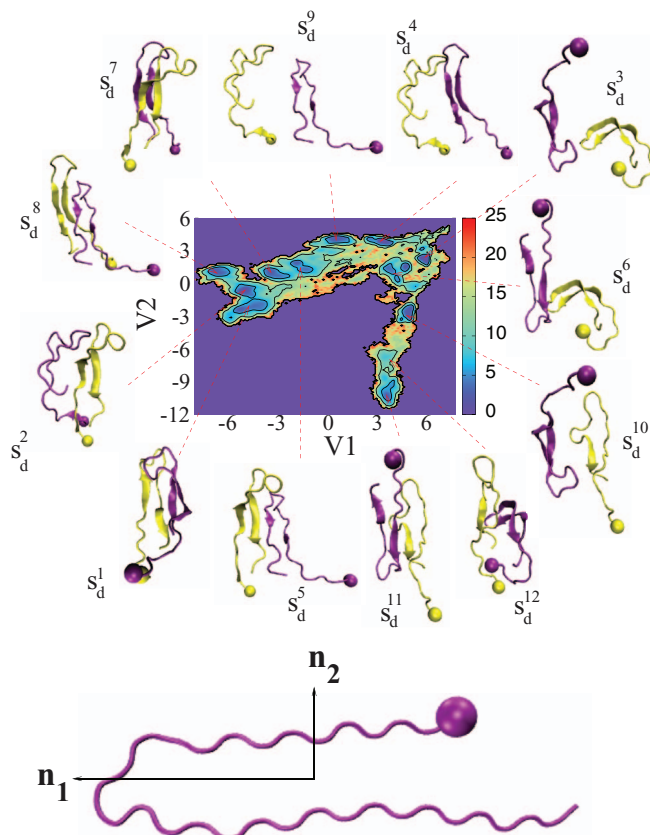| Single-molecule state					$(A\beta_{9-40})_4$ Intramolecular state		
s_s^k	P	β	Turn	Coil	S_{intra}^m	P	Combination $\{s_s^k\}$
1	36	35	31	34	1	33	5 4 1 1
2	23	28	28	44	2	24	3 3 2 1
3	19	40	30	30	3	20	3 2 2 1
4	12	23	47	30	4	4	4 3 1 1
5	9	26	32	42	5	4	4 2 2 1

FIG. 4. Free energy landscape (in kcal/mol) of the double-molecule states of Aβ_{9-40} projected on the first two principal components V1 and V2 obtained from PCA. The center of each metastable double-molecule state is also shown. The ball indicates the first residue. The two axis vectors ($\mathbf{n_1}$, $\mathbf{n_2}$) laying in the plane of a monomer are just defined to describe the orientation of the chains.

The product basis of these five single-molecule states, s_s^k, yields to $5^4 = 625$ possible combinations for the intramolecular states, S_{intra}^m ($m = 1, \ldots, 625$), of (Aβ_{9-40})$_4$. After automatic removal of the degenerate states, and because many

combinations are not sampled during the simulation, we obtain 37 unique combinations. Of these, the first five states, covering 85% of the total population, are given in Table I, with their description in terms of the single-molecule states s_s^k. It is found that all five states S_{intra}^m ($m = 1, \ldots, 5$) of (Aβ_{9-40})$_4$ contain the single-molecule state s_s^1, and only the first state S_{intra}^1 contains the fifth state s_s^5.

B. Intermolecular states of (Aβ_{9-40})$_4$

We now characterize the intermolecular states of (Aβ_{9-40})$_4$ in terms of the structures of double-molecules. Each conformation of the tetramer consists of $6 \times 32 \times 32 = 6144$ distances between the inter-chain side-chain's centers of mass. These distances completely specify the intermolecular structures of a conformation of the tetramer. Now, we map all these 6144 distances to $32 \times 32 = 1024$ distances: $d_i^m(t_n) \mapsto D^m(t_j)$ ($m = 1, \ldots, 1024$, $i = 1, \ldots, 6$), which are considered as distances sampled by a virtual double-chain simulated alone at times $t_j = t_n, \ldots, t_n + 5$. We then apply the PCA of the inverse distances $1/D^m$ to obtain the most dominant structures. We construct the FEL along the first relevant two principal components, which account for about 73% of the system's fluctuation.

The FEL shown in Fig. 4 reveals the existence of 12 *double-molecule states* s_d^k. Again, these states are expected to be different from those sampled by the dimer simulated alone. The conformation of each double-chain of Aβ_{9-40} can switch between these 12 states. Table II characterizes the structures of these states by the average distance between the inter-chain centers of mass $\langle D \rangle$, the mean number of inter-chain backbone hydrogen bonds $\langle N_{\text{hb}} \rangle$, the mean number of inter-chain side-chain contacts $\langle N_c \rangle$, and the orientation between two chains. To describe the orientation between the two chains, we define for each monomer k two principal axis vectors ($\mathbf{n_1^k}$, $\mathbf{n_2^k}$) laying in the plane of the monomer [Fig. 4]. Overall, there are five orientational arrangements:

TABLE II. (Left) Structural characterization of the double-molecule states s_d^k ($k = 1, \ldots, 12$), shown in Fig. 4, with their populations P (in %), distances between the centers of mass D (in nm), orientations of the two vectors ($\mathbf{n_1^k}$, $\mathbf{n_2^k}$) for each monomer k, and the total number of inter-peptide H-bonds (N_{hb}) and inter-peptide side-chain contacts (N_c). (Right) The intermolecular states S_{inter}^l ($l = 1, \ldots, 12$) of (Aβ_{9-40})$_4$ with their populations P (in %) formed by the combinations $\{s_d^k\}$ of the double-molecule states.

		Double-molecule states					(Aβ_{9-40})$_4$ Intermolecular states	
s_d^k	P	D	Orientation	N_{hb}	N_c	S_{inter}^l	P	Combination $\{s_d^k\}$
1	13.3	0.67	$\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}, \mathbf{n_2^1} \uparrow\uparrow \mathbf{n_2^2}$	7	39	1	40.9	9 8 7 5 4 2
2	11.4	0.99	$\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}, \mathbf{n_2^1} \perp \mathbf{n_2^2}$	4	27	2	27.4	11 10 6 3 1 1
3	8.8	2.39	$\mathbf{n_1^1} \perp \mathbf{n_1^2}, \mathbf{n_2^1} \perp \mathbf{n_2^2}$	0	0	3	11.8	11 10 6 3 2 1
4	8.7	1.65	$\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}, \mathbf{n_2^1} \perp \mathbf{n_2^2}$	2	14	4	5.6	12 11 6 3 1 1
5	8.5	1.50	$\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}, \mathbf{n_2^1} \uparrow\downarrow \mathbf{n_2^2}$	7	20	5	3.4	9 8 7 7 4 2
6	8.4	1.88	$\mathbf{n_1^1} \perp \mathbf{n_1^2}, \mathbf{n_2^1} \perp \mathbf{n_2^2}$	1	10	6	2.8	8 7 5 5 4 2
7	8.2	1.09	$\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}, \mathbf{n_2^1} \uparrow\downarrow \mathbf{n_2^2}$	1	23	7	2.6	9 8 5 5 4 2
8	8.0	1.02	$\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}, \mathbf{n_2^1} \uparrow\uparrow \mathbf{n_2^2}$	7	35	8	2.2	12 6 6 3 2 1
9	7.8	2.32	$\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}, \mathbf{n_2^1} \perp \mathbf{n_2^2}$	0	0	9	1.8	12 10 6 3 1 1
10	7.3	1.47	$\mathbf{n_1^1} \uparrow\downarrow \mathbf{n_1^2}, \mathbf{n_2^1} \uparrow\uparrow \mathbf{n_2^2}$	0	4	10	1.3	8 5 5 5 4 2
11	7.2	0.95	$\mathbf{n_1^1} \uparrow\downarrow \mathbf{n_1^2}, \mathbf{n_2^1} \uparrow\uparrow \mathbf{n_2^2}$	9	34	11	1.3	12 11 6 3 2 1
12	1.8	0.97	$\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}, \mathbf{n_2^1} \uparrow\uparrow \mathbf{n_2^2}$	1	28	12	1.2	9 8 7 7 5 4

(1) Fully parallel configurations with $\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}$ and $\mathbf{n_2^1} \uparrow\uparrow \mathbf{n_2^2}$; states s_d^1 : 13%, s_d^8 : 8%, and s_d^{12} : 2%. The two chains are close in space with the N-terminus (C-terminus) of one chain making contacts with the N-terminus (C-terminus) of another chain. These states are different however in the values of the inter-chain distances and the number of H-bond and contacts [Table I].

(2) Parallel/perpendicular configurations with $\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}$ and $\mathbf{n_2^1} \perp \mathbf{n_2^2}$; states s_d^2: 11%, s_d^4: 9%, s_d^9: 8%. The two peptides have an overall parallel orientation in terms of n1, but the β-hairpin of one chain is perpendicular to the β-hairpin of another chain, and the states differ in the inter-chain distances, and the numbers of hydrogen bonds and contacts.

(3) Fully perpendicular configurations with $\mathbf{n_1^1} \perp \mathbf{n_1^2}$ and $\mathbf{n_2^1} \perp \mathbf{n_2^2}$; states s_d^3: 9%, s_d^6: 8%. These configurations are characterized by large separations between the two chains, $\langle D \rangle = 2.4$ and 1.9 nm.

(4) Parallel/antiparallel configurations with $\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}$ and $\mathbf{n_2^1} \uparrow\downarrow \mathbf{n_2^2}$; states s_d^5: 8.5%, s_d^7: 8.2%. The N-terminus (C-terminus) of one chain is in proximity with the C-terminus (N-terminus) of another chain.

(5) Antiparallel/parallel configurations with $\mathbf{n_1^1} \uparrow\downarrow \mathbf{n_1^2}$ and $\mathbf{n_2^1} \uparrow\uparrow \mathbf{n_2^2}$: states s_d^{10}: 7.3%, s_d^{11}: 7.2%. In these configurations, the two peptides are antiparallel. Although these two states have similar orientations, the chains in s_d^{10} have zero inter-chain H-bonds and four inter-chain side chain contacts, while the chains in s_d^{11} form many H-bonds and contacts.

Overall, the analysis shows that the double-molecule states are different in their orientations, positions, and internal structures. Having enumerated 12 double-molecule states, s_d^k, we now identify the intermolecular states of $(A\beta_{9-40})_4$. In principle, there are $12^6 = 2.985.984$ combinations that 6 double chains can be distributed into 12 double-molecule states. After automatic removal of the degenerate states and because many combinations are not sampled during the simulation, we obtain 48 unique intermolecular states, S_{inter}^l ($l = 1, \ldots, 48$), for $(A\beta_{9-40})_4$. Of these, the first 12 states, covering 95% of the total population, are listed in Table II, together with their combinations of double-molecule states, s_d^k, that form the intermolecular states of the tetramer.

We can classify the intermolecular structures into two groups based on the orientations of the chains. The first group consists of six states S_{inter}^l, $l = 1, 5, 6, 7, 10, 12$, which cover 52% of the total population. These states are formed by double-molecule states s_d^k, whose dimers form parallel-like configurations ($\mathbf{n_1^1} \uparrow\uparrow \mathbf{n_1^2}$), and the second vectors $\mathbf{n_2}$ can be parallel, antiparallel, or perpendicular. Their compositions can differ by one or more double-molecule states. For example, the state S_{inter}^1 is defined by the combination {9, 8, 7, 5, 4, 2}, while the state S_{inter}^5 is defined by the combination {9, 8, 7, 7, 4, 2} and the state S_{inter}^{12} is defined by the combination {9, 8, 7, 7, 5, 4}. The second group consists of six states S_{inter}^l, $l = 2, 3, 4, 8, 9, 11$, which cover 48% of the total population. These states are formed by double-molecule states with dimers in parallel, antiparallel, or perpendicular orientations [see Fig. 4]. They can differ from each other by one to four double-molecule states.

C. Overall conformational states of $(A\beta_{9-40})_4$

Having identified the intermolecular, S_{inter}^l, and intramolecular, S_{intra}^m, states, we now combine these information to obtain the full 3D picture of the overall conformational states, $S_{overall}^n$, of $(A\beta_{9-40})_4$. The overall conformational states are simply the product basis of 48 S_{inter}^l ($l = 1, \ldots, 48$) states and 37 S_{intra}^m ($m = 1, \ldots, 37$) states. Of these $48 \times 37 = 1776$ combination states, only 133 unique combinations are sampled by the simulation, and the first 9 states account for almost 90% of the full population. The remaining states are only explored by a few MD snapshots, and thus can be safely discarded.

The nine overall states are described in Table III in terms of their product basis ($S_{inter}^l \times S_{intra}^m$), population, total numbers of inter-peptide H-bonds and side-chain contacts, secondary structure contents, and numbers of p-stranded β-sheets. Representative structures of these nine states are shown in Fig. 5. These nine states have a β-strand content varying between 28% and 36%, a turn content varying between 29% and 35%, and a coil content varying between 32%

TABLE III. Structural characterization of the overall states, $S_{overall}^n$ ($n = 1, \ldots, 9$), of $(A\beta_{9-40})_4$, including the product basis of the intermolecular states S_{inter}^l and intramolecular states S_{intra}^m ($l \times m$), their populations P (in %), the total number of the inter-peptide H-bonds (N_{hb}) and side-chain contacts (N_c), the secondary structure contents (in %), and the number of p-stranded β-sheets ($N_{ps}, p = 2, \ldots, 6$). N_{2s} includes intramolecular states (β-hairpins) and intermolecular states.

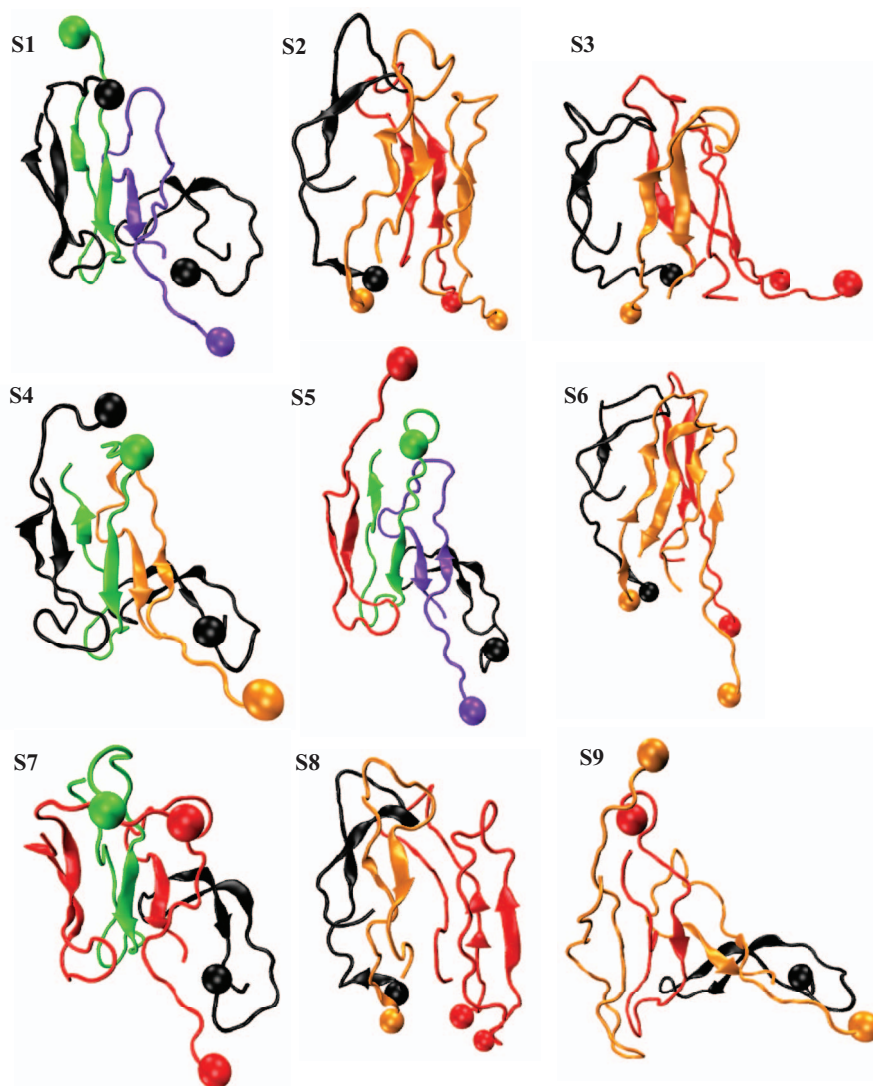| $S_{overall}^n$	$S_{inter}^l \times S_{intra}^m$	P	N_{hb}	N_c	β	Turn	Coil	N_{2s}	N_{3s}	N_{4s}	N_{5s}	N_{6s}
1	2×1	36.6	19	117	30	35	35	2	0	0	0	0
2	1×2	22.5	20	123	36	30	34	3	0	0	0	0
3	1×3	19.1	18	127	32	29	38	3	0	0	0	0
4	2×4	4.0	20	119	33	35	32	0	0	0	0	1
5	2×7	4.4	16	106	28	34	38	1	0	0	1	0
6	5×2	2.3	24	142	36	30	34	2	1	0	0	0
7	8×5	2.0	16	90	28	33	39	2	0	0	0	0
8	7×3	1.0	20	112	33	29	38	3	0	0	0	0
9	6×2	1.0	22	97	36	30	34	2	0	0	0	0

FIG. 5. The representative structures of $(A\beta_{9-40})_4$ belonging to the overall states shown in Table III. Chains are colored according to their single-molecule states on the free energy landscape shown in Fig. 3, whose secondary structure contents are given in Table I. The ball indicates the first residue.

and 38%. The number of interpeptide H-bonds varies between 16 and 24, and the number of interpeptide side-chain contacts varies between 90 and 142.

As seen from $S_{inter}^l \times S_{intra}^m$, the overall states can differ in the intramolecular state, the intermolecular state or in both states. The conformational structures of $S_{overall}^1$, $S_{overall}^4$, and $S_{overall}^5$ representing 45% of all structures, have the same intermolecular state S_{inter}^2, characterized by mixed parallel/antiparallel chains. They differ, however, in the intramolecular structure of the four peptides. The most populated tetrameric state $S_{overall}^1$ (37%), is characterized by S_{intra}^1, that is the combination {5, 4, 1, 1} of the intramolecular state of each single-molecule (see Table I), while $S_{overall}^4$ (4%) is characterized by S_{intra}^4 and the combination {4, 3, 1, 1}.

The state $S_{overall}^1$ is characterized by one 2-stranded β-sheet formed by residues 18–21 (green chain) and 34–37 (blue chain), and a β-hairpin formed by residues 16–22, 30–37 (left black chain). The state $S_{overall}^4$ contains one 6-stranded β-sheet formed by consecutive strands spanning residues: 18–23, 31–36 (right black chain), 16–19, 34–37 (orange chain), 18–21

(green chain), and 16–19 (left black chain). The state $S_{overall}^5$ displays one β-hairpin spanning residues 16–21, 31–36 (red chain), and one 5-stranded β-sheet formed by consecutive strands spanning residues: 18–22, 32–36 (right black chain), 16–19, 34–37 (blue chain), and 18–21 (green chain). The state $S_{overall}^7$ with a population of 2.0% has also the same mixed parallel/antiparallel orientation of the peptides, and is characterized by two β-hairpins spanning residues 16–19 and 34–36 (left red chain) and 18–23 and 31–36 (black chain). This state has a much lower number of interpeptide contacts (Nc = 90) than the $S_{overall}^1$, $S_{overall}^4$, and $S_{overall}^5$ states (N_c = 117, 119, and 106, respectively).

Peptides in the states $S_{overall}^2$, $S_{overall}^3$, $S_{overall}^6$, and $S_{overall}^8$, representing 45.9% of all structures, have similar orientations along the main axis vectors $\mathbf{n_1}$, i.e., the four chains are in parallel, but the orientations along the second axis vectors $\mathbf{n_2}$ are different. For example, the vector $\mathbf{n_2}$ of one orange chain is perpendicular to that of the red chain in $S_{overall}^2$, while these two vectors tend to be parallel in $S_{overall}^6$, resulting in a 3-stranded β-sheet with a high number of contact N_c = 142

[Table III]. The state S^2_{overall} displays two β-hairpins spanning residues 18–20 and 32–34 (left orange chain) and 14–15 and 33–36 (red chain), and one 2-stranded β-sheet spanning residues 11–14 (left orange chain) and 11–15 (black chain). The state S^3_{overall} is characterized by three β-hairpins spanning residues 14–15 and 31–35 (left red chain), 16–20 and 32–33 (orange chain), and 16–20 and 35–36 (black chain). The state S^6_{overall} contains one β-hairpin formed by residues 14–22 and 29–35 (red chain), one 2-stranded β-sheet spanning residues 11–12 (left orange chain) and 11–12 (black chain), and one 3-stranded β-sheet formed by consecutive strands spanning residues 16–20 and 30–39 (left orange chain) and 28–32 (right orange chain). The state S^8_{overall} contains one β-hairpin spanning residues 18–20 and 32–34 (orange chain), one 2-stranded β-sheet spanning residues 11–12 (orange chain) and 11–12 (black chain), and another 2-stranded β-sheet spanning residues 14–17 (left red chain) and 18–21 (right red chain). In terms of the intramolecular structure, S^2_{overall} and S^6_{overall} are β-rich (25% and 33%), whereas S^3_{overall} and S^8_{overall} have β contents of 20% and 16%, respectively. Finally, the state S^9_{overall} is quite disordered in both intermolecular structure ($N_c = 97$) and intramolecular structure (only 17% β). It is composed of two β-hairpins spanning residues 16–17 and 34–37 (right orange chain) and residues 18–23 and 31–36 (black chain).

It is of interest to compare our results with previous simulations and experiments. Though our MD simulations are too short to explore equilibrium ensemble, they provide strong evidence that the Aβ_{9-40} tetramer can form, in addition to parallel chains with a high percentage of β-hairpins and various chain-to-chain angles, long-lived mixed parallel/antiparallel chains with various conformations of the peptides and chain-to-chain angles. Atomistic MD simulations on the Aβ_{9-40} dimer showing that structures with an antiparallel β-sheet between the two CHC regions remaining stable within 200 ns at 310 K (in preparation), four antiparallel CHC regions, not sampled in this 400 ns simulation, are also long-lived states for the tetramer.

So, in contrast to MD simulations of Aβ_{1-42} and Aβ_{9-42} trimers starting from the perfect fibrillar states and remaining in a parallel β-sheet structure,[33,34] our all-atom simulations indicate that the population of fully parallel chains with intermolecular stranded β-sheets is likely to be very marginal for the tetramer. Three simulations using implicit solvent models support our findings. Klimov[53] found that the tetramer of Aβ_{9-40} lacks parallel β-sheets, though their all-atom REMD simulations predict much higher α helix content than the experiments. Coarse-grained replica-exchange molecular dynamics simulations of the trimer of Aβ_{17-42}, and coarse-grained DMD simulations of Aβ_{1-42} dimers and tetramers, did not report any parallel β-sheet structure.[54,55]

We also note that although the impact of the N-terminus on Aβ oligomerization cannot be ignored,[2,9,11,20,47,55–57] our high percentage of β-hairpins and thus antiparallel β-sheet structures in tetramers is supported by FTIR and time-resolved hydrogen exchange mass spectrometry experiments on Aβ_{1-40} and Aβ_{1-42} oligomers.[58,59]

In summary, our simulations indicate the existence of many long-lived non-fibrillar tetramer topologies with high turn-coil content and various β-strands of various lengths and positions (including the residues 11–15) adding further complexity to our understanding of the formation of Aβ fibrils at the molecular level.

IV. CONCLUSIONS

We have proposed a new method allowing the characterization of the structures of amyloid oligomers having arbitrary shapes and position-orientation arrangements. The method takes into account both the intermolecular and intramolecular degrees of freedom, and treats correctly the degenerate states. The method consists of three main steps and gives the following results for the tetramer (A$\beta_{9-40})_4$:

- The intermolecular structures of an oligomer are described in terms of the combination of double-molecule states. These states describe the position-orientation arrangements of a double-chain in the presence of the other double-chains, and can be obtained through the PCA method of the inverse distances between side-chain's centers of mass of two chains of the double-molecule trajectory. For (A$\beta_{9-40})_4$, we found 12 double-molecule states with different orientations, e.g., parallel, antiparallel, perpendicular between two chains. The combination of these 12 states yields to 48 unique intermolecular states for (A$\beta_{9-40})_4$, and 12 states include 95% of the ensemble.

- The intramolecular structures of an oligomer are described in terms of the combination of single-molecule states. These states describe the structures of a single-chain in the presence of the other chains, and can be obtained through the dPCA method of the single-molecule trajectory. For (A$\beta_{9-40})_4$, we found 5 dominant single-molecule states with different conformational properties. The combination of these 5 states results in 37 unique intramolecular states for (A$\beta_{9-40})_4$. Of these, 5 states represent 85% of the total population.

- The overall structure of the tetramer is given in terms of the product basis of the intermolecular and intramolecular structures. In total, we obtained 133 unique overall structures for (A$\beta_{9-40})_4$ and 9 states are sufficient to represent 90% of all generated structures.

While step 2 was proposed by Riccardi et al.,[32] step 1 is new. It allows us to characterize the intermolecular structures of amyloid systems having arbitrary shapes and adopting any position-orientation arrangements. This step is in contrast with the classification of the intermolecular structures proposed in Ref. 32 which is system-dependent and only suitable for molecules adopting extended conformations.

Our automatic method can be applied to any oligomer size for two reasons. First, the identification of the single-molecule and double-molecule states involves the PCA of single- and double-molecule trajectories, and this is straightforward. In this work, we found that it is sufficient to use only the first two principal components as they always cover more than 60% of the system's fluctuations. Including more components into the analysis is, however, easy by employing the k-means clustering analysis.[60,61] Second, while the number

of combinations of the single-molecule or double-molecule states increases with the number of chains, especially for highly flexible molecules, the calculation of the combinations is simple and fast, and the identification of the overall structures in step 3 is, again, a simple matter. In addition, many states can be marginally populated and safely discarded.

Currently, we are using this method to characterize the conformational states of the oligomer formed by 16 Aβ_{37-42} peptides from a long atomistic REMD simulation in explicit solvent (800 ns/replica).

[1] D. J. Selkoe, Science **337**, 1488 (2012).
[2] Z. Lv, R. Roychaudhuri, M. M. Condron, D. B. Teplow, and Y. L. Lyubchenko, Sci. Rep. **3**, 2880 (2013).
[3] S. L. Bernstein *et al.*, Nat. Chem. **1**, 326 (2009).
[4] J. E. Straub and D. Thirumalai, Curr. Opin. Struct. Biol. **20**, 187 (2010).
[5] G. Wei, W. Song, P. Derreumaux, and N. Mousseau, Front. Biosci. **13**, 5681 (2008).
[6] A. De Simone and P. Derreumaux, J. Chem. Phys. **132**, 165103 (2010).
[7] Y. Chebaro and P. Derreumaux, Proteins **75**, 442 (2009).
[8] J. Nasica-Labouze, M. Meli, P. Derreumaux, G. Colombo, and N. Mousseau, PLoS Comput. Biol. **7**, e1002051 (2011).
[9] S. Cote, R. Laghaei, P. Derreumaux, and N. Mousseau, J. Phys. Chem. B **116**, 4043 (2012).
[10] J. E. Shea and B. Urbanc, Curr. Top. Med. Chem. **12**, 2596 (2012).
[11] T. Zhang, J. Zhang, P. Derreumaux, and Y. Mu, J. Phys. Chem. B **117**, 3993 (2013).
[12] Y. Lu, P. Derreumaux, Z. Guo, N. Mousseau, and G. Wei, Proteins **75**, 954 (2009).
[13] F. Baftizadeh, X. Biarnes, F. Pietrucci, F. Affinito, and A. Laio, J. Am. Chem. Soc. **134**, 3886 (2012).
[14] R. Pellarin, P. Schuetz, E. Guarnera, and A. Caflisch, J. Am. Chem. Soc. **132**, 14960 (2010).
[15] G. Bellesia and J. E. Shea, J. Chem. Phys. **131**, 111102 (2009).
[16] V. A. Wagoner, M. Cheon, I. Chang, and C. K. Hall, J. Mol. Biol. **416**, 598 (2012).
[17] S. Auer, C. Dobson, and M. Vendruscolo, HFSP J. **1**, 137 (2007).
[18] A. De Simone *et al.*, Proc. Natl. Acad. Sci. U.S.A. **109**, 6951 (2012).
[19] B. Ma and R. Nussinov, J. Biol. Chem. **286**, 34244 (2011).
[20] T. Zhang, W. Xu, Y. Mu, and P. Derreumaux, ACS Chem. Neurosci. **5**(2), 148–159 (2014).
[21] P. Nguyen, M. S. Li, J. E. Staub, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **104**, 111 (2007).
[22] P. H. Nguyen and P. Derreumaux, J. Phys. Chem. B **117**, 5831 (2013).
[23] W. Song, G. Wei, N. Mousseau, and P. Derreumaux, J. Phys. Chem. B **112**, 4410 (2008).
[24] D. Matthes, V. Gapsys, and B. L. de Groot, J. Mol. Biol. **421**, 390 (2012).
[25] Y. Mu, P. H. Nguyen, and G. Stock, Proteins **58**, 45 (2005).
[26] W. Xu, C. Zhang, L. A. Morozova-Roche, J. Z. Zhang, and Y. Mu, J. Phys. Chem. B **117**, 8392 (2013).
[27] H. B. Nam, M. Kouza, Z. Hoang, and M. S. Li, J. Chem. Phys. **132**, 165104 (2010).
[28] W. Xu, C. Zhang, P. Derreumaux, A. Grslund, L. Morozova-Roche, and Y. Mu, PLoS ONE **6**, e24329 (2011).
[29] G. Reddy, J. E. Straub, and D. Thirumalai, J. Phys. Chem. B **113**, 1162 (2009).
[30] P. H. Nguyen, M. S. Li, and P. Derreumaux, Phys. Chem. Chem. Phys. **13**, 9778 (2011).
[31] G. Stock, A. Jain, L. Riccardi, and P. H. Nguyen, "Exploring the free energy landscape of small peptides and proteins by molecular dynamics simulations," in *Protein and Peptide Folding, Misfolding and Non-folding*, edited by R. Schweitzer-Stenner (John Wiley and Sons, Inc., Hoboken, NJ, USA, 2011).
[32] L. Riccardi, P. H. Nguyen, and G. Stock, J. Chem. Theory. Comput. **8**, 1471 (2012).
[33] M. F. Masman *et al.*, J. Phys. Chem. B **113**, 11710 (2009).
[34] A. H. C. Horn and H. Sticht, J. Phys. Chem. B **114**, 2219 (2010).
[35] W. van Gunsteren *et al.*, *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag AG an der ETH, Zurich, 1996).
[36] H. J. C. Berendsen, J. Postma, W. van Gunsteren, and J. Hermans, *Intermolecular Forces* (Reidel, Dortrecht, 1996).
[37] A. T. Petkova *et al.*, Proc. Natl. Acad. Sci. U.S.A. **99**, 16742 (2002).
[38] H. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[39] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Mod. **7**, 306 (2001).
[40] J. P. Ryckaert, G. Cicotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[41] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[42] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[43] D. Frishman and P. Argos, Proteins **23**, 566 (1995).
[44] J. A. Hartigan and M. A. Wong, Appl. Stat. **28**, 100 (1979).
[45] *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2008).
[46] "Nbclust: An examination of indices for determining the number of clusters," 2013, see http://cran.r-project.org/web/packages/nbclust.
[47] M. H. Viet, P. H. Nguyen, S. T. Ngo, M. S. Li, and P. Derreumaux, ACS Chem. Neurosci. **4**, 1446 (2013).
[48] T. Ichiye and M. Karplus, Proteins **11**, 205 (1991).
[49] A. Kitao, F. Hirata, and N. Go, Chem. Phys. **158**, 447 (1991).
[50] A. E. García, Phys. Rev. Lett. **68**, 2696 (1992).
[51] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, Proteins **17**, 412 (1993).
[52] P. L. Clark, Trends Biochem. Sci. **29**, 527 (2004).
[53] S. R. Shaikh, Biophys. J. **99**, 1 (2010).
[54] Y. Chebaro *et al.*, J. Phys. Chem. B **116**, 8412 (2012).
[55] B. Urbanc, M. Betnel, L. Cruz, G. Bitan, and D. B. Teplow, J. Am. Chem. Soc. **132**, 4266 (2010).
[56] P. H. Phuong, B. Tarus, and P. Derreumaux, J. Phys. Chem. B **118**, 501 (2014).
[57] P. H. Nguyen and P. Derreumaux, Acc. Chem. Res. **47**(2), 603–611 (2014).
[58] J. Pan, J. Han, C. Borchers, and L. Konermann, Biochemistry **51**, 3694–3703 (2012).
[59] E. Cerf *et al.*, Biochem. J. **421**, 415 (2009).
[60] A. Altis, M. Otten, P. H. Nguyen, R. Hegger, and G. Stock, J. Chem. Phys. **128**, 245102 (2008).
[61] A. Altis, P. H. Nguyen, R. Hegger, and G. Stock, J. Chem. Phys. **126**, 244111 (2007).