Anatomy and Formation Mechanisms of Early Amyloid-β Oligomers with Lateral Branching: Graph Network Analysis on Large-Scale Simulations

Miao Yuan,[†] Xuan Tang,[†] Wei Han[*†]

†State Key Laboratory of Chemical Oncogenomics, Guangdong Provincial Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, 518055, China

*Corresponding author: email: hanw@pkusz.edu.cn

Supplementary Information

Calculation of the Flory's characteristic ratio

The Flory's characteristic ratio C_n can be used to measure the stiffness of the oligomer, which can be calculated as

$$C_n = \frac{\langle R^2 \rangle}{nr^2}, \text{ [S1]}$$

where r is the average distance between neighboring peptides, n is the number of edges in the longest shortest path of the oligomer, and $\langle R^2 \rangle$ is the mean square end-to-end distance of the path. The corresponding values for Ab40 and Ab42 were calculated to be 20.8 Å and 20.5 Å, respectively. C_n can vary with *n* when *n* is small but will reach the value C_∞ that is unchanged at large *n* values, as shown in **Fig. S5b**. If C_∞=1, the oligomer resembles a freely jointed chain. For C_∞=2~3, the oligomer behaves like a freely rotating chain, and an even larger C_∞ indicates that the oligomer could be more extended due to hindered internal bond rotation.

Secondary structure content analysis

The secondary structure analysis of Aβ peptides was mainly based on backbone dihedral angles (ϕ, ψ) and the formation of hydrogen bonds between amino acid residues. Residue k is thought to be in an α-helical state[1] if the backbone dihedral angles (ϕ, ψ) of this residue and its two nearest neighbors in sequence are all within (-60±30°, -47±30°). Any two residues are thought to form β-sheet structure if their backbone dihedral angles are both within (-135±45°, 135±45°) and their adjacent amide groups form HB interactions with one another. Finally, for any four consecutive residues in sequence, if the distance between the C_α carbon atoms of the first and fourth residues is shorter than 7 Å, the two middle residues are thought to form a turn given that they are not a part of any helical structure.[2]

Conformational clustering into tertiary structural states

To examine global conformational transition of Aβ, we first sought to coarse-grain conformational space into large basins, each comprised of many conformations sharing a similar tertiary topology. To this end, we clustered conformations using an approach

proposed by Meng et al..[3] In this approach, each conformation is represented with a contact probability matrix $\boldsymbol{M}$ in which its element m_{ij} indicates how likely residues *i* and *j* of a peptide forms contact. A contact is formed if the two residues' minimum atomic distance is shorter than 4.5 Å. Of note, we did not consider the contacts formed between residues separated by no more than four residues in sequence. The distance d_{AB} between two conformations *A* and *B* is defined as

$$d_{AB} = 1 - \frac{\frac{1}{w^2} + 2\sum_{i=1}^{w}\sum_{j=1}^{w} m_{ij}^{A} m_{ij}^{B}}{\frac{1}{w^2} + \sum_{i'=1}^{w}\sum_{j'=1}^{w}\left(m_{i'j'}^{A} + m_{i'j'}^{B}\right)} \quad , \qquad [S2]$$

where M_A and M_B are respective contact probability matrices of the two conformations and *w* is the number of residues in the peptide. Based on this distance metrics, the k-mean algorithm was employed to cluster conformations. As our goal here is to identify distinct regions of tertiary conformational space accessible for Aβ chains, we used the sampled conformations of all peptide chains contained in the simulations regardless of their assembly status. For each peptide chain, its conformation was recorded every 13.5 ns as a time-averaged contact probability matrix over this duration.

In the k-mean clustering method, the number of clusters *N* is a hyper-parameter that needs to be pre-defined. To find a minimum value for this parameter that still ensures the quality of clustering results, we assessed a quantity[3] called L_2 which measures the distance between the average contact matrix over all conformations and the weighted average contact matrix of the central conformations of resulting clusters with the weight of each cluster proportional to its population. The L_2 distance decreases as the cluster number *N* increases, but the change in L_2 will become less significant beyond certain values of *N* which can be used as a reasonable estimate of the minimum number of clusters needed for clustering (**Fig. S4a** and **Fig. S4b**). Following the previous approach,[3] we assessed the variation of L_2 with respect to the increment at *N* using a quantity Δ_N defined as

$$\Delta_N = \frac{|L_2(N) - L_2(N-1)|}{L_2(2)} \quad . \qquad [S3]$$

Similar to the previous study,[3] a Δ_N cut-off of 10% was imposed. **Fig. S4c** and **Fig. S4d** shows Δ_N calculated using the clustering results for Aβ40 and Aβ42 with $N \in$

[2,12]. In both cases, Δ_N was above 10% until N = 4, suggesting that at least four tertiary structural states are needed to represent the conformations of both alloforms.

Estimation of time scale of conformational rearrangement

To probe the conformational dynamics of Aβ in oligomer states, we analyzed the simulation period $t \in [5\mu s, 15\mu s]$ during which most of Aβ chains participated in the formation of oligomers. We assumed that the conformational rearrangement of each Aβ chain was independent to the others, and a single simulation can thus be thought to sample 100 independent instances of Aβ structural dynamics in an average oligomeric environment. An Aβ chain was thought to undergo global structural transition if this peptide was observed to sample at least two different tertiary structural states with the probability of each state > 20%. As these global transition events are infrequent, they can be treated as a two-state Poisson process. The mean first passage time of the structural transition, τ, of Aβ can be written as,[4,5]

$$\tau = -\frac{Ht}{\ln(1-\lambda)} \quad , \quad [S4]$$

where $t = 10\mu s$ is the length of the simulation period used for analysis, H = 200 is the total number of independent instances of Aβ included in the two simulations, and λ is the fraction of instances that were observed to undergo structural transition, which was calculated to be 0.03 and 0.05 for Aβ40 and Aβ42, respectively.

Table S1. Percentage secondary structure contents of Aβ in oligomers of different sizes (n).

	Aβ40			Aβ42		
	$2 \leq n \leq 3$	$4 \leq n \leq 10$	$n > 10$	$2 \leq n \leq 3$	$4 \leq n \leq 10$	$n > 10$
α-helix	1.8	2	2.1	1.6	1.5	1.6
β-sheet	15.8	14.5	13.5	15.8	16.2	15.8
Turn	43	44.5	43.8	47.2	46.8	46.6

Table S2. Comparison of cross collision-section areas measured experimentally and calculated in this study

	Aβ40 CCS (Å^2)			Aβ42 CCS (Å^2)			Aβ42 - Aβ40 CCS (Å^2)	
n/z	Expt.[a]	Sim.	Diff.[c]	Expt.[a]	Sim.	Diff.	Expt.	Sim.
1/-3	620	695±7[b]	75	702	705±8	3	82	10
2/-5	1142	1138±1	-4	1256	1154±6	-102	114	16
4/-10	2080	2020±10	-60	2332	2100±44	-232	252	80
6/-15		2949±95		2898	3061±44	163		112

a) Obtained from IM-MS experiment.[6] b) Standard errors were determined as absolute difference in average CCS values between the two trajectories for each case. c) Deviation of calculated CCS values from experimental ones.

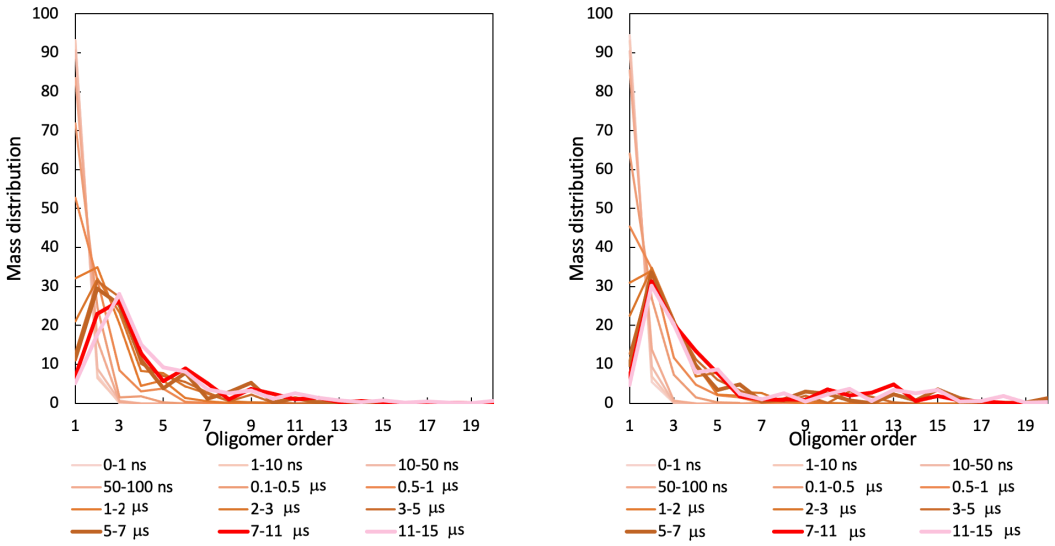

Figure S1. Time evolution of oligomer mass distributions of Aβ40 (left) and Aβ42 (right). The curves are averaged over two independent simulations for both Aβ40 and Aβ42.

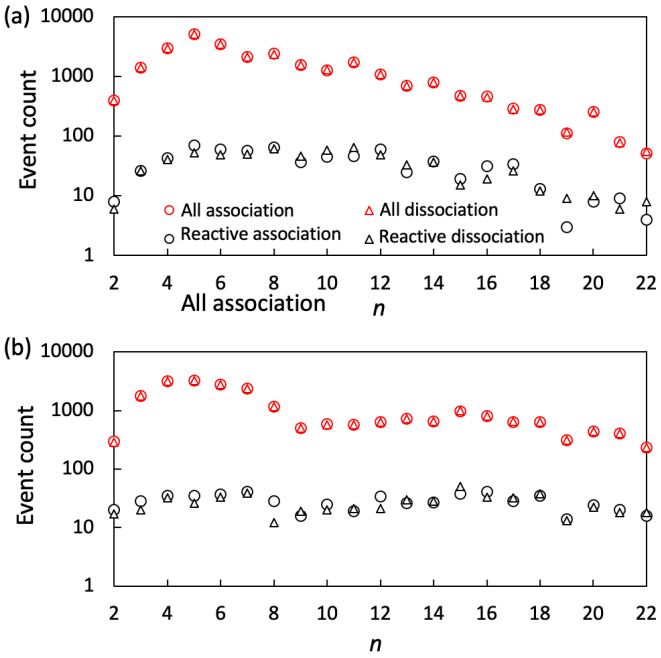

Figure S2. Event counts of formation and fragmentation of Aβ40 (a) and Aβ42 (b) oligomers of different sizes n. All the events were observed during the simulation period $t \in [5\ \mu s, 15\ \mu s]$. Circles and triangles indicate the counts of association and dissociation events, respectively. Red symbols denote the counts of all the events observed and black symbols denote the counts of those reactive ones.

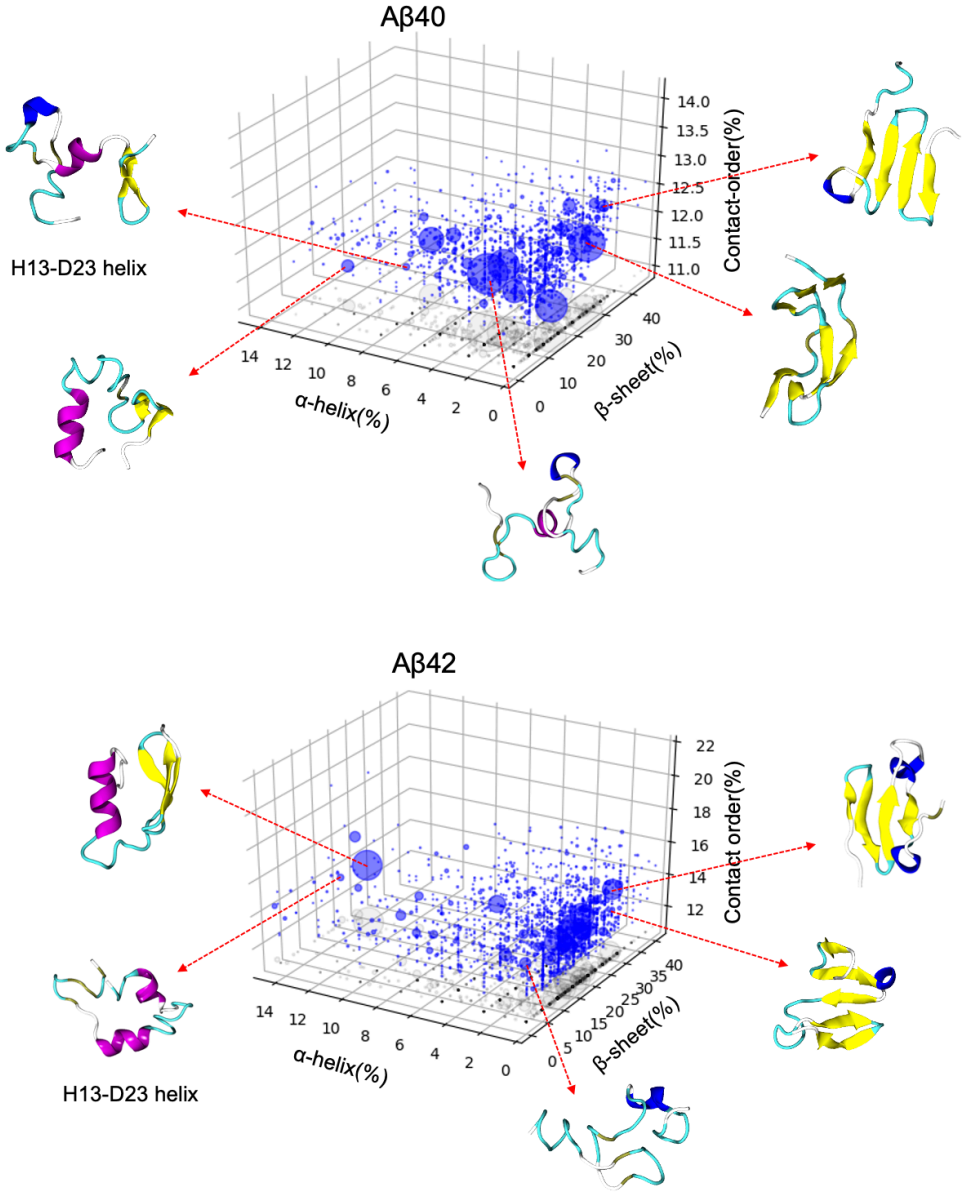

Figure S3. Conformational ensembles of Aβ40 (top) and Aβ42 (bottom) monomers sampled in simulations. For each alloform, all its conformational clusters are plotted as spheres in a 3D space whose *x*-, *y*- and *z*-axes denote the average α-helical, β-sheet and contact order of clusters, respectively. The contact order was calculated as $1/L \cdot N \sum_1^N \Delta L_{ij}$, where N is the number of atomic contacts, ΔL_{ij} is the number of residues separating a contact between atoms i and j, and L is the length of peptides in amino acid residues. The clusters were obtained through conformational clustering of sampled structures based on all-atom RMSD with a 2 Å RMSD cutoff. The size of each sphere is proportional to the population of the corresponding cluster. Grey shades denote the projection of clusters on the (α%, β%) plane. Color coding in representative structures: coil (white), turn (cyan), β-sheet (yellow), α-helix (purple), and 3_{10}-helix (blue).

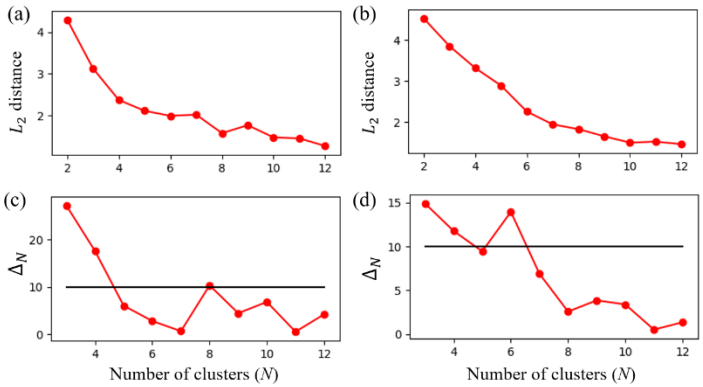

Fig S4. Conformations of Aβ in oligomers can be classified into four tertiary structural states. (a,b) The L_2 distances obtained after the conformations of Aβ40 (a) and Aβ42 (b) were clustered into the different numbers of states. (c,d) Incremental improvement Δ_N of quality of clustering results . The cut-off of Δ_N used to determine the minimum cluster numbers is shown in the black horizontal line.

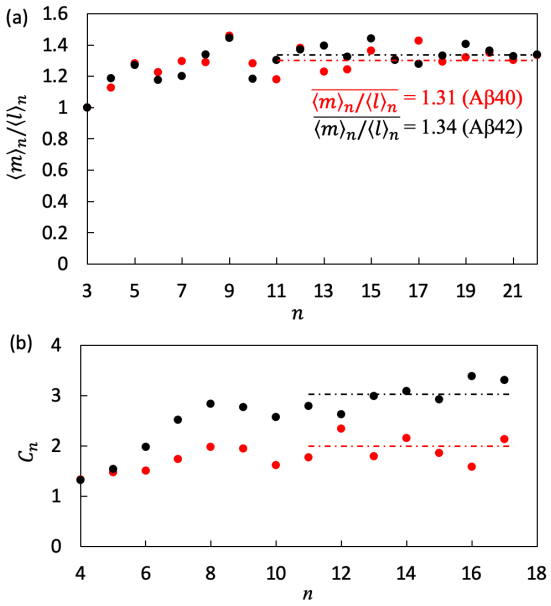

Figure S5. Average thickness (a) and extendedness (b) of Aβ40 (red dots) and Aβ42 (black dots) oligomers. The oligomer thickness was calculated as the ratio of the average number of backbone nodes ($\langle m\rangle_n$) to the average length of the longest shortest path ($\langle l\rangle_n$). The oligomer extendedness was estimated using the Flory's characteristic ratio C_n. Both properties vary with small n values but, when $n > 10$, they fluctuate around the mean values indicated by the dashed dotted lines.

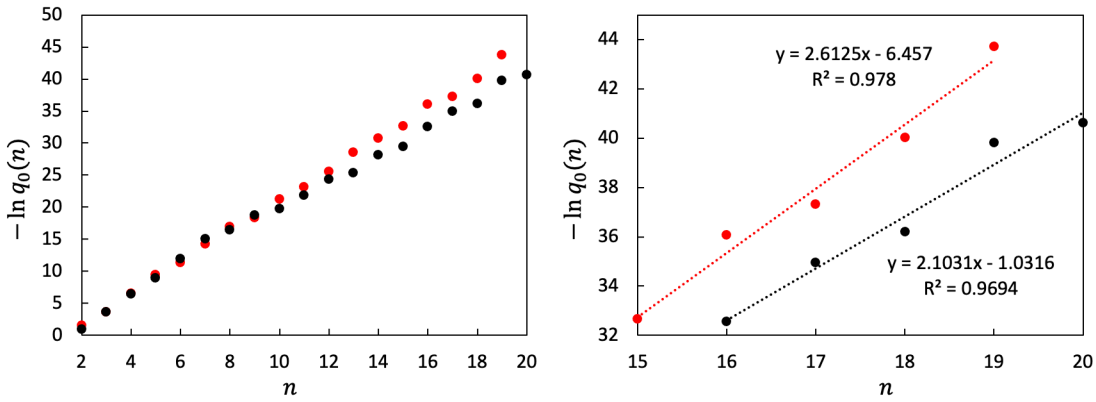

Figure S6. Plots of $-\ln q_0(n)$ against oligomer size n. Red and black dots denote the results for Aβ40 and Aβ42 oligomers, respectively. The panel on the right shows the linear correlation between $-\ln q_0(n)$ and n for $n \geq 15$, as indicated by the dotted lines.

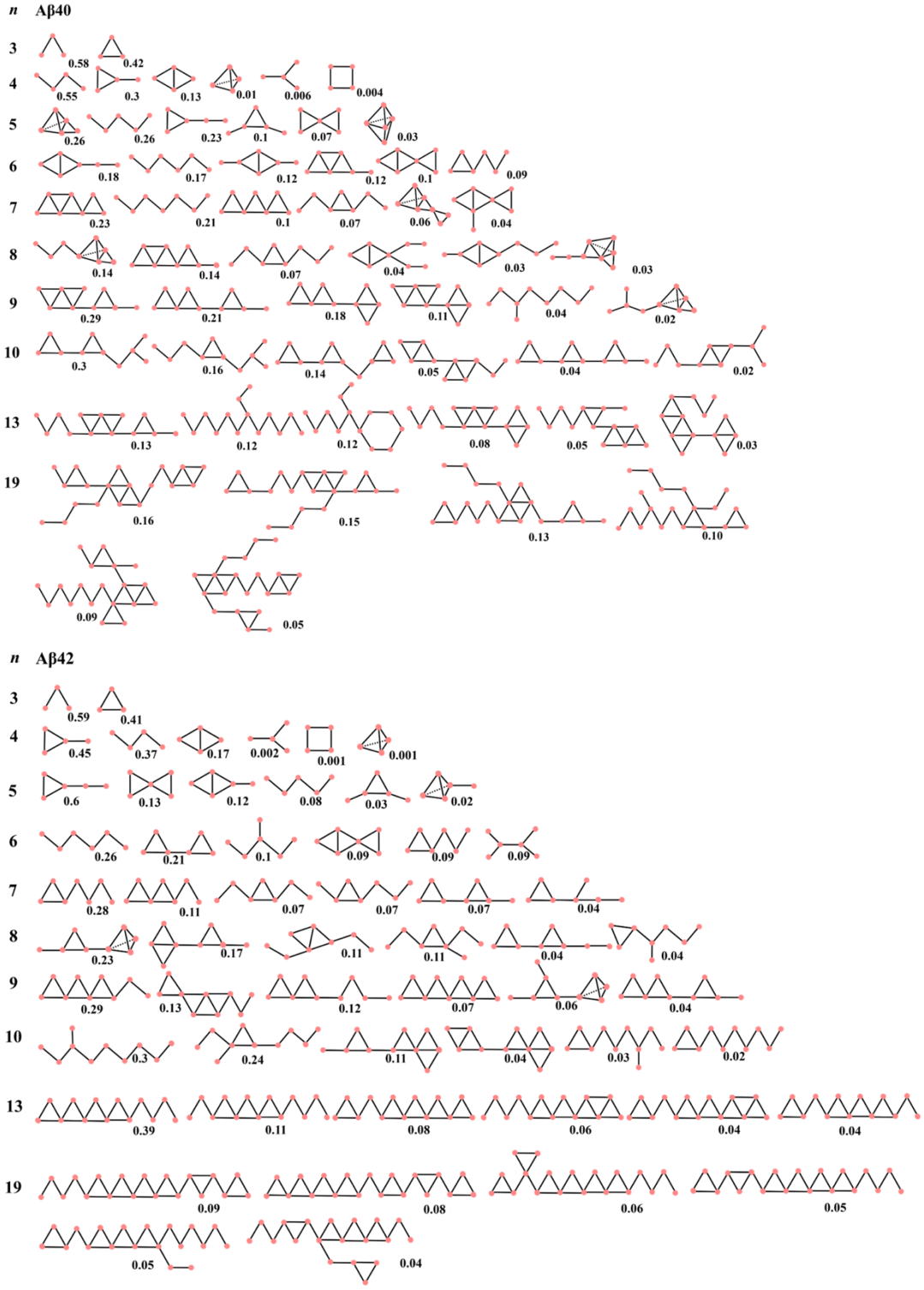

Figure S7. The most popular topologies of Aβ40 (top) and Aβ42 (bottom) oligomers identified with our clustering algorithm. Orange dots represent Aβ chains in oligomers and black line segments denote the physical contacts between these chains. Shown below each topology is its corresponding fractional population.

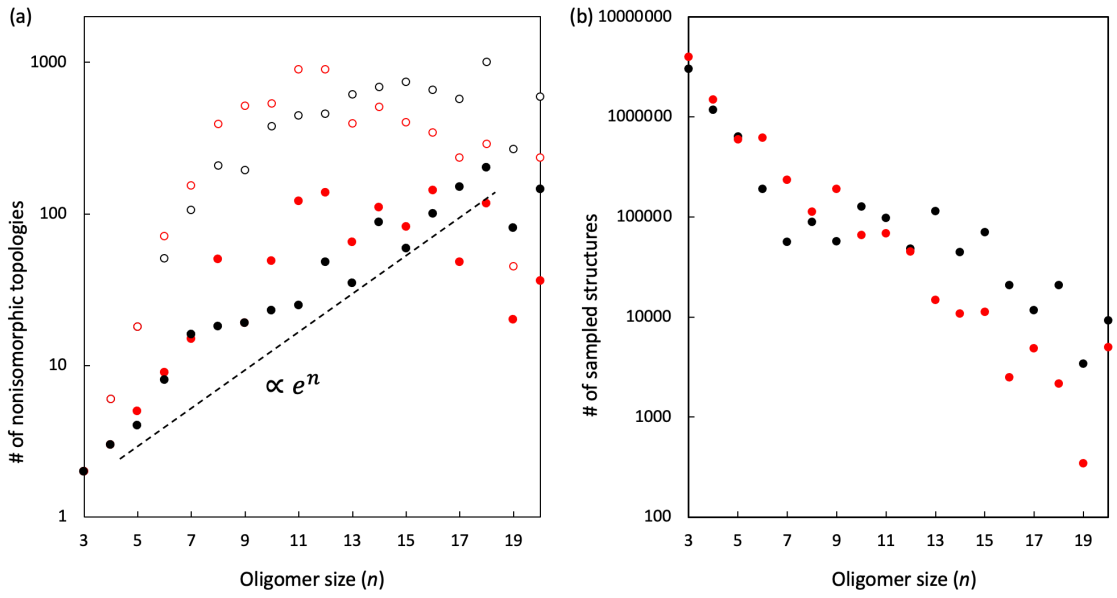

Figure S8. Extensive topological sampling of our simulations. (a) The numbers of non-isomorphic topologies (circles) of Aβ oligomers of size n and those most probable ones (dots) (with a combined probability > 90%) identified in our simulations. (b) The number of oligomer structures sampled in our simulations. The results for Aβ40 and Aβ42 oligomers are shown in red and black, respectively. The dashed line in (a) is for the guide of eyes, showing the exponential growth of populated topologies with size *n*. This trend continues until $n > 18$ where the converged structural sampling becomes more difficult.

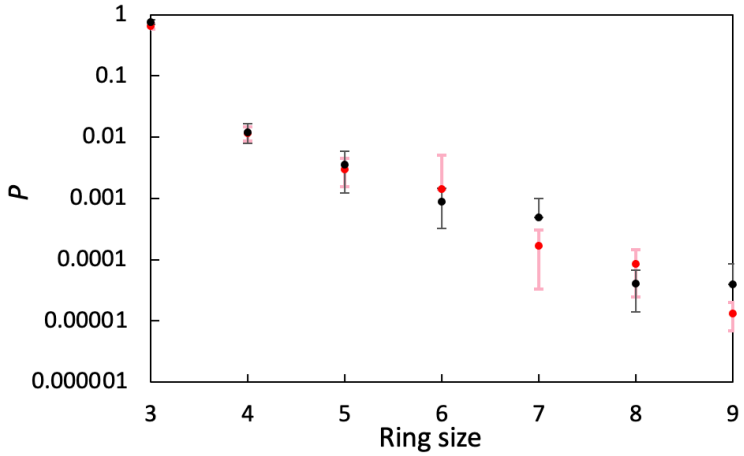

Figure S9. Chance of observing at least a ring structure of a given size in Aβ40 (red dots) and Aβ42 (black dots) oligomers of size $n \geq 4$.

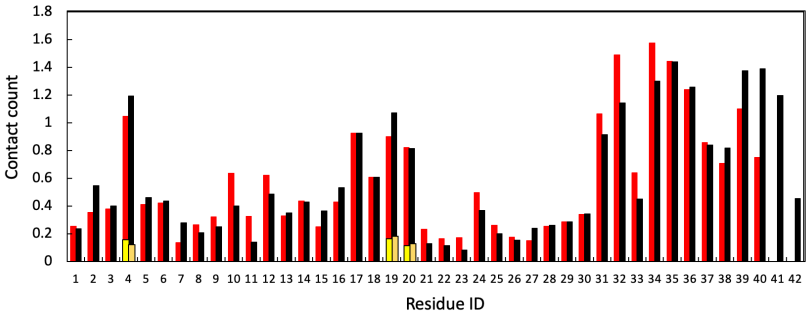

Figure S10. The average count of residual contacts formed between a given residue of an Aβ peptide in an oligomer and any other residues of the peptide's neighbors. Red and Black bars denote the results for Aβ40 and Aβ42, respectively. Yellow bars denote the chance of a given aromatic residue forming intermolecular aromatic-aromatic contacts in oligomers.

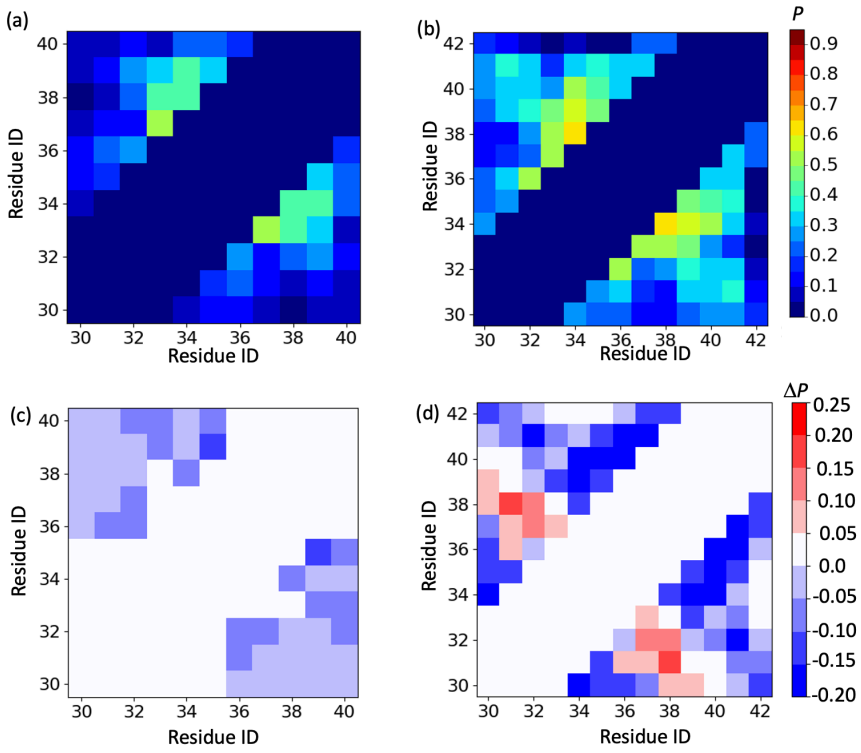

Figure S11. Intramolecular contact probability in the C-terminal regions of (a) Aβ40 and (b) Aβ42. The change of intramolecular contact probability in the same regions of (c) Aβ40 and (d) Aβ42 upon the formation of at least 5 intermolecular residual contacts between the C-terminal regions at branching interfaces. The color bar indicates the scale of probability values.

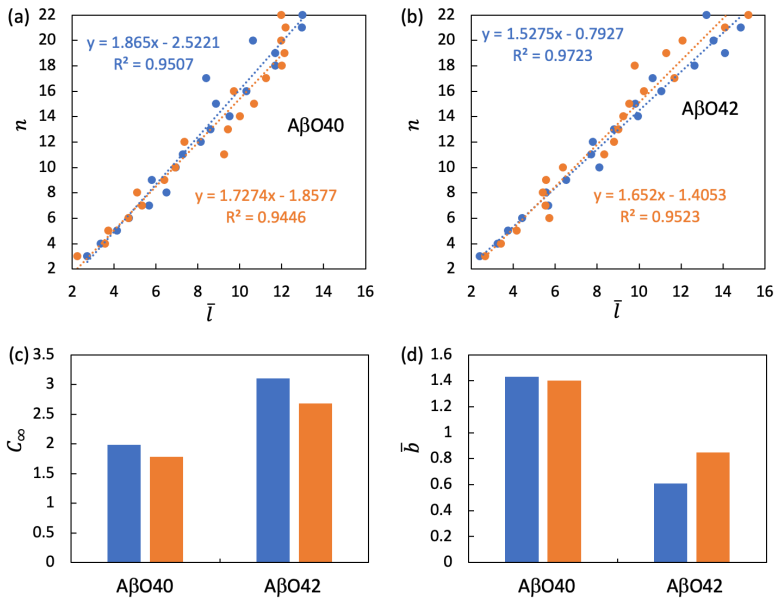

Figure S12. Dependence of simulation results on initial monomer conformations used. Plots of oligomer size n against average backbone length $\bar{l}$ for (a) Aβ40 and (b) Aβ42 oligomers. (c) The Flory's characteristic ratio C_∞ of AβOs with $n > 10$. (d) Average branch count $\bar{b}$ of AβOs with $n > 15$. For each type of oligomers, two independent simulations were performed, starting from different initial monomer conformations. The results obtained from the first and the second simulation are shown in blue and orange, respectively. All the results are obtained using $t \in [10\mu s, 15\mu s]$ of each simulation trajectory.

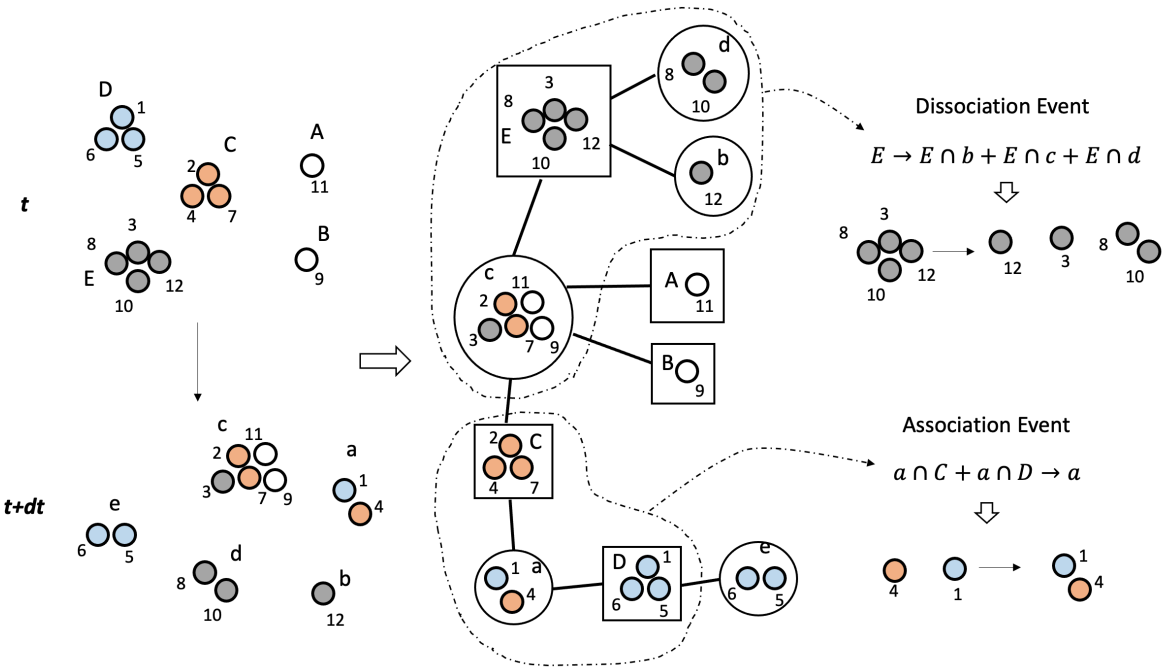

Figure S13. Scheme of workflow of association/dissociation event detection algorithm. Two timeframes at *t* and *t* + *dt* on the left are the examples used to demonstrate how the algorithm works. Each peptide is indicated by a circle and numbered and all the circles in the same cluster at *t* are colored the same. The clusters at *t* and *t* + *dt* are labeled "A, B, …" and "a, b, …", respectively. The middle panel illustrates the auxiliary network representation of the two timeframes that is then used to identify association/dissociation events as shown in the right panel. Of note, the example shown here is a very special case that is used purposely to show the robustness of our algorithm. In our simulations, the situation like this rarely occurred.

REFERENCES

1 E. J. Sorin and V. S. Pande, *Biophys J*, 2005, **88**, 2472-2493.

2 C. M. Wilmot and J. M. Thornton, *J Mol Biol*, 1988, **203**, 221-232.

3 F. Meng, M. M. J. Bellaiche, J. Y. Kim, G. H. Zerze, R. B. Best and H. S. Chung, *Biophys J*, 2018, **114**, 870-884.

4 V. A. Voelz, G. R. Bowman, K. Beauchamp and V. S. Pande, *J Am Chem Soc*, 2010, **132**, 1526-1528.

5 H. Yu, W. Han, W. Ma and K. Schulten, *J Chem Phys*, 2015, **143**, 243142.

6 S. L. Bernstein, N. F. Dupuis, N. D. Lazo, T. Wyttenbach, M. M. Condron, G. Bitan, D. B. Teplow, J. E. Shea, B. T. Ruotolo, C. V. Robinson and M. T. Bowers, *Nat Chem*, 2009, **1**, 326-331.