

Supplementary Information: Dynamics of the Conformational Transitions during the Dimerization of an Intrinsically Disordered Peptide: a Case Study on the Human Islet Amyloid Polypeptide Fragment

Qin Qiao^{1*}, Ruxi Qi², Guanghong Wei^{2*}, Xuhui Huang^{3,4,5*}

¹Hefei National Laboratory for Physical Sciences at the Microscale and Collaborative Innovation Center of Chemistry for Energy Materials (iChEM), University of Science and Technology of China, Hefei, 230026, China

²State Key Laboratory of Surface Physics, Key Laboratory for Computational Physical Sciences (MOE), and Department of Physics, Fudan University, Shanghai, China

³Department of Chemistry, ⁴Division of Biomedical Engineering, ⁵Center of Systems Biology and Human Health, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Email: qinqiao@connect.ust.hk; ghwei@fudan.edu.cn; xuhuihuang@ust.hk

This supplementary document includes five sections: 1. the validation of the Markov State Models; 2. the detailed structures of the top 6 most populated metastable states; 3. the time evolution of the minimum distance starting from the uncollapsed chains; 4. the structural details of the top 4th to 6th initial collapsed states; 5. the free energy and the detailed structures of the microstates along the top 6 pathways from the initial collapsed states to the most populated microstate featured with antiparallel cross- β structure; 6. Structural comparison between the fully-extended- β state and the final state along the 6th pathway.

1. Model validation

Implied time scales

The implied time-scales indicate the relaxation time of each transition mode in the model. When the model becomes Markovian, the implied time-scales should be constant with the increase of the lag time. For the microstate model, the implied time-scales are shown in SI Fig. 1a. There exist a bundle of slow transitions at time-scales around hundreds of μ s, as well as the faster transitions at tens of μ s. In comparison, the implied time-scales for the 33-state macrostate model are shown in SI Fig. 2a. The slowest time-scale is around hundreds of μ s, consistent with that of the microstate model. Moreover, in the macrostate model, there is a large separation between the slowest and the other slow time-scales, which are around several μ s.

The implied time-scales of both the microstate and macrostate model level off from lag-time equal to 30ns, indicating 30ns is the time interval required for the system dynamics to be Markovian.

Residence probability tests

When being Markovian, the model built at short lag time can be propagated to obtain the long-time dynamics. To test whether 30 ns is long enough for the system dynamics to be Markovian, we compared the state residence probability obtained from the MSMs with that from the MD trajectories as a function of time. As shown in SI Fig. 1b, for the microstate model, the residence probability from model prediction (red curves) agrees with that from the real trajectories (black curves) for the most populated microstates. Similarly, the two curves agree with each other in the macrostate model, as shown in SI Fig. 2b.

2. Detailed structures of metastable states

In addition to the top 3 most populated macrostates discussed in the main-text, the 4th to 6th most populated macrostates are also included here. As shown in SI Fig. 4a, in the 4th most populated macrostate occupying 7.1% population, the Chain-1 (blue colored) adopts the β -hairpin structure, while the Chain-2 (green colored) has coiled structures. The inter-molecular contacts are at the N-termini via hydrophobic interactions. The 5th macrostate (5.3%) is mainly coiled, and has some bend and turn structures. The inter-molecular contacts are at C-termini via the hydrogen bonds. Last but not least, in the 6th macrostate (4.4%), the two chains intersect at the middle region, and the N-terminus of the blue-colored Chain-1 form antiparallel inter-molecular contacts with the C-terminus of the green-colored Chain-2. In comparison with the top 3 most populated ones, the 4th-6th most macrostates have substantial intra-molecular interactions, and the structures are more collapsed.

The most populated microstates in the top 6 most populated macrostates are also shown in SI Fig. 5-10. Though flexible in some parts of the structures, the overall structural arrangements have similar feature in each macrostate.

3. Minimum distance between two chains during the initial collapsing processes

The time-scales of initial collapsing processes are shorter than the Markovian lag time (30 ns). As shown in SI Fig. 11, the time-dependent inter-molecular minimum distance distribution indicates

the two chains collapse rapidly. In detail, the minimum distance has already decreased to 2.5 Å in the 5-to-10 ns time window (red curve), deviating from that in the initial 5 ns starting from separate chains (black curve). Furthermore, the two chains remain collapsed afterwards, as shown in the rest curves in SI Fig. 11. This observation indicates the two chains collapse together rapidly (~10 ns), and the probability for the initial collapsed complexes to dissociate right away is negligible.

4. Detailed structures of initial collapsed states

In addition to the top 3 initial collapsed states discussed in the main-text, we also analyzed the 4th to 6th most probable initial collapsed states. In the 4th initial collapsed state, the two chains form antiparallel hydrogen bonds with each other at their C-termini, as shown in SI Fig. 12a. In addition, the middle region of the blue-colored Chain-1 is highly bended.

In contrast, the initial inter-molecular contacts in 5th initial collapsed state are mainly between residue ¹¹R and ¹⁵F through π - π interaction. Both the two chains are highly bended in the middle region, and there are also some β -bridge secondary structures as shown in SI Fig. 12b.

In the 6th initial collapsed state, the two chains form helical structures, and the two helices associate side-by-side in antiparallel orientation. This antiparallel association is energetically favorable due to the electrostatic dipole interactions.

5. Detailed pathways from initial collapsed states to the most populated antiparallel cross- β structure

As mentioned in the main-text, the slow time-scales of the transitions from initial collapsed states to the most populated antiparallel cross- β structure indicate that the dimer free energy landscape is rugged. To demonstrate the ruggedness, we plot the free energy of each microstate along the top 6 pathways, as shown in SI Fig. 13. The free energy f_a of state a is obtained via $f_a = -kT \ln(p_a)$, where p_a is the equilibrium population of state a from MSMs. As shown in SI Fig. 13, the free energy difference between the initial collapsed states and the final state is around $3 kT$, while the barriers along the pathways range from 1.1 to 3.4 kT approximately. The overlap between these two free energy differences indicates there exist metastable intermediates on the pathways and the free energy landscape is rugged.

The representative pathways initialized from the top 6 initial collapsed states are plotted in SI Fig. 14-19. Except the 1st pathway, the other pathways contain multiple intermediates. Moreover, the 2nd to the 6th pathways go through some expanded coiled structures, and transit through the various metastable states, from initial collapsing states to the final antiparallel cross- β state.

Contrary to the ‘one-step nucleation’ mechanism, the antiparallel cross- β dimer state is formed through the conformational rearrangements under the interplay between both the intra- and inter-molecular interactions, i.e., there exist further structural rearrangements during the antiparallel cross- β dimer formation. To examine the cooperativity of the conformational changes of each chain, we calculated the change of RMSDs between the intermediate states and the final antiparallel cross- β state, and the number of residue-residue contacts, along the top 6 pathways. The RMSDs are calculated for each chain separately, i.e., the rotation and translation is performed based on each single chain, instead of the whole dimer. An inter-molecular contact is formed when the minimal distance between two residues is smaller than the cutoff, which is equal to 3.85 Å. An intra-chain contact is formed based on the same cutoff, provided that the two residues are separate by at least two residues in the sequence. As shown in the SI Fig. 20, the RMSDs of each chain between the intermediate states and the final antiparallel cross- β state decrease in the similar trend along the top 6 pathways. In the meanwhile, the number of intra-molecular contacts also decreases together along each pathway, as shown in SI Fig. 21. These results indicate that the conformational change of each chain happens cooperatively, instead of sequentially.

6. Structural comparison between the fully-extended- β state and the final state along the top 6th pathway.

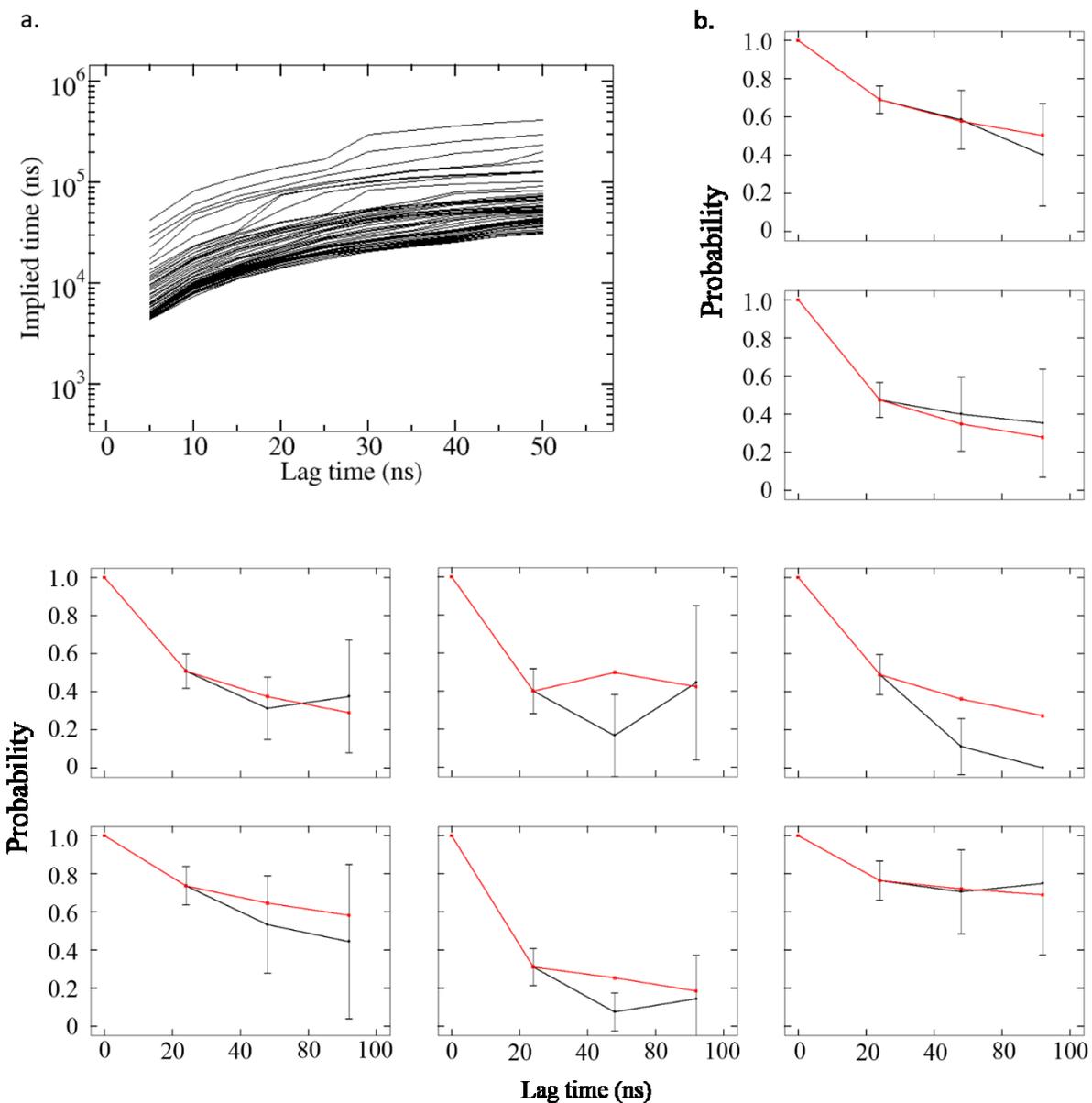
The 6th pathway starting from α -helix structures passes through an intermediate state featured with more extended β structures than the final state. The forward committor probability p^+ of this fully-extended- β state is equal to 0.80. Before reaching the final state, this fully-extended- β state is followed by another intermediate state. These three microstates all belong to the most populated macrostate featured with antiparallel cross- β structures. To compare the structural difference among these three states, we carried out further analysis as follows.

First, as shown in the SI Fig. 22, the probability of the β -sheet structure decreases from the fully-extended- β state to the final state. In the fully-extended- β state, the inter-chain β -sheet structures persist throughout residue 13-22 in both the two chains, as shown in SI Fig. 22(a). In the

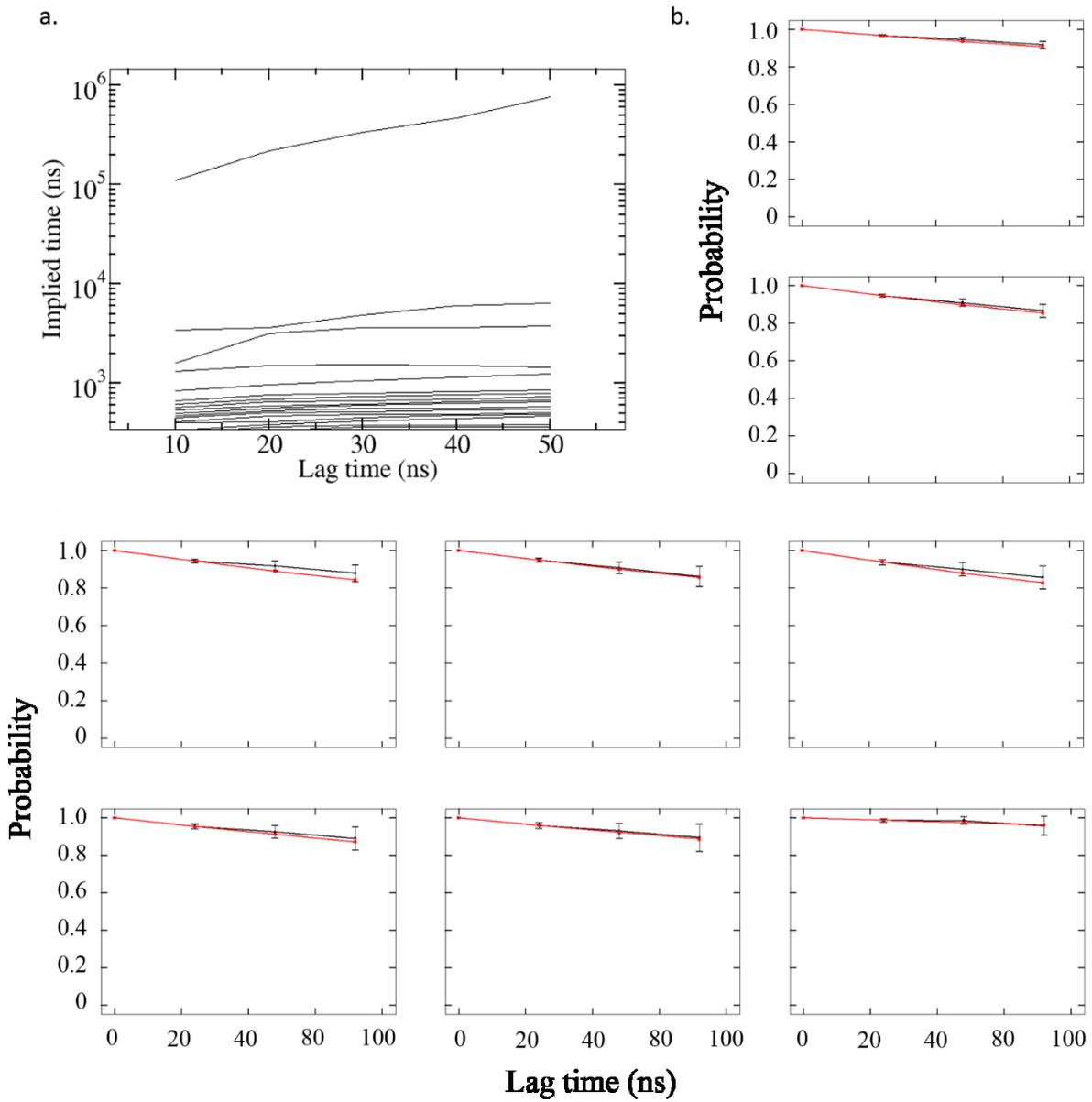
intermediate state shown in the SI Fig. 22(b), the middle residues (¹⁶L in Chain-1, ¹⁸H in Chain-2) begin to adopt bend structures. These bended structures in the middle region become more obvious and partially change to coils in the final state, where there exist two antiparallel inter-chain β -sheet fragments. The short fragment is formed between residue 13-14 in Chain-1 and residue 22-23 in Chain-2, while the long one between residue 19-24 in Chain-1 and residue 12-17 in Chain-2. Thus, we label these three states as ‘Extended- β ’, ‘Bended- β ’ and ‘Coiled- β ’ respectively. In all the three states, the two chains associate with each other in antiparallel direction.

Despite their difference in the secondary structure, the residue-based contact maps of these three states are very similar. As shown in SI Fig. 23, all the three states are featured with antiparallel inter-molecular contacts. The difference among these three states lies in the short (~ 2 -residue-long) inter-molecular parallel contacts in the middle region of the peptide, which appears only in the ‘Bended- β ’ intermediate and ‘Coiled- β ’ final states.

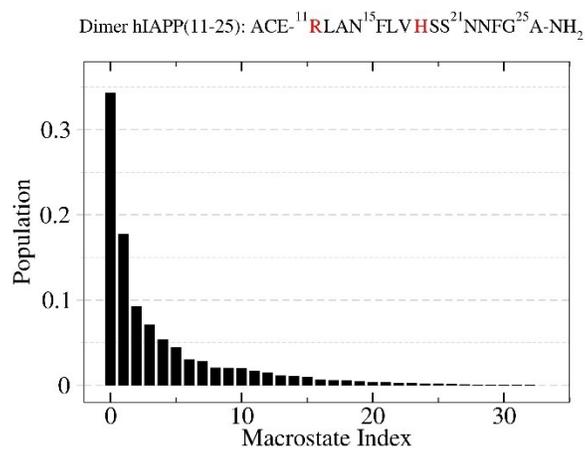
The ‘Extended- β ’ state is less stable than the ‘Coiled- β ’ final state thermodynamically, despite the more extended inter-chain β -sheet structures in the former. As shown in the SI Fig. 24(a), the free energy of the ‘Coiled- β ’ state is around $1.5-kT$ lower than that of the ‘Extended- β ’ state. Besides, the ‘Bended- β ’ state has the highest free energy among these three states, which is around $1.0-kT$ higher than that of the ‘Extended- β ’. The origin of the low free energy of the final state may arise from its collapsed geometry, which prevents the full explosion of the hydrophobic residues to the solvent. As shown in the SI Fig. 24(b), the radius of gyration of the hydrophobic residues decreases significantly from the ‘Extended- β ’ state to the ‘Coiled- β ’. In the meanwhile, however, the number of inter-chain hydrogen bonds only decreases slightly, as shown in SI Fig. 24(c). This marginal decrease of hydrogen bonds might be surprising at first sight, as the ‘Extended- β ’ state has more ordered inter-chain β -sheet structures. However, as shown in SI Fig. 22(c), there are some newly-formed inter-molecular hydrogen bonds between the sidechains in the middle region (residues ¹⁸HSSN²²N) in the final state, which compensate energetically for the loss of the main-chain hydrogen bonds.



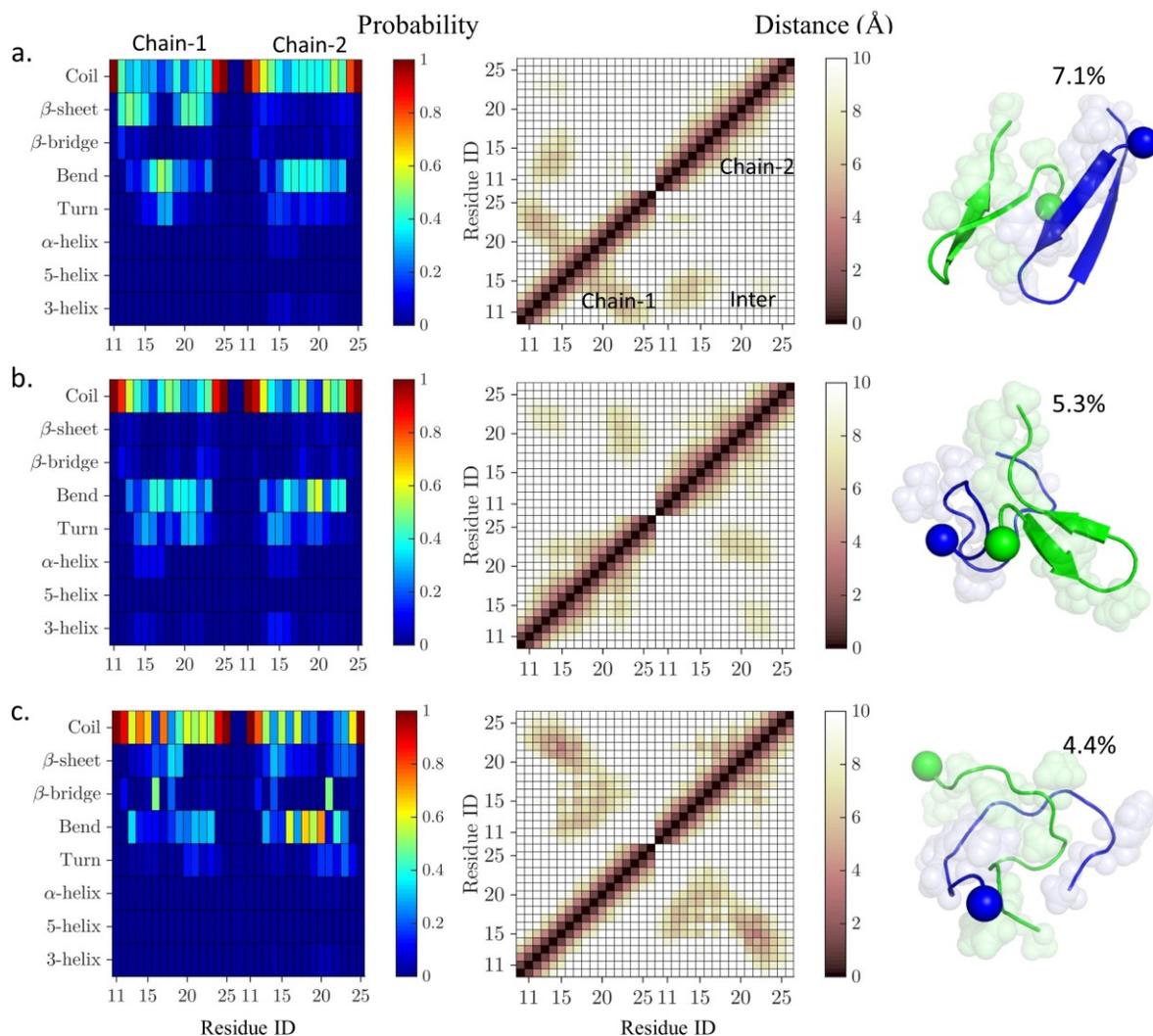
SI Fig. 1 (a) Implied times-scales plot of 12,000-state microstate model. The implied time-scales level off from 30ns. The longest implied time reaches hundreds of microseconds. (b) Residence probability test with the lag time equal 30ns for the most populated 8 microstates. The black curve is obtained from MD simulations and the red one from the propagation of MSMs.



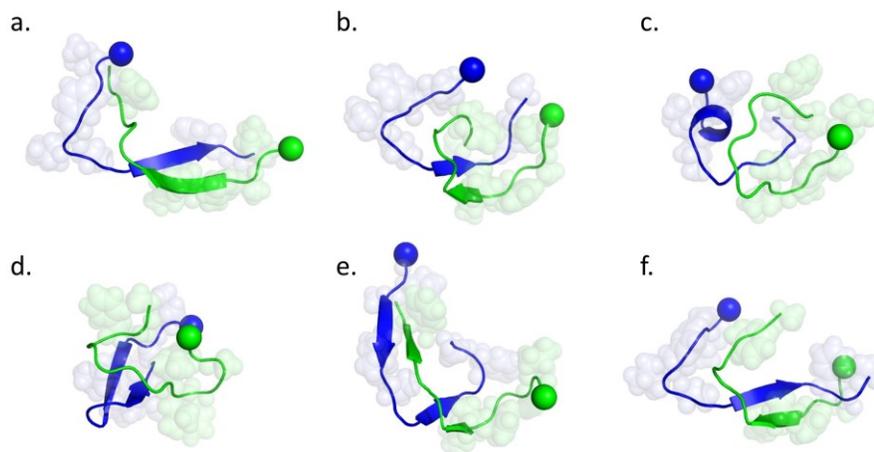
SI Fig. 2 (a) Implied timescales plot for the 33-state macrostate model (b) Residence probability test for the top 8 most populated macrostates with lag time equal 30ns. The black curve is obtained from MD simulations and the red one from the propagation of MSMs.



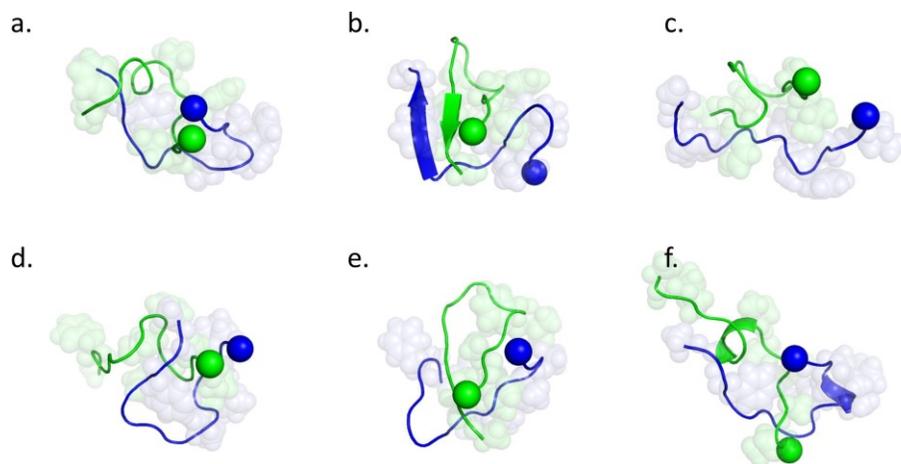
SI Fig. 3 The sorted population distribution of the 33-state macrostate model. The most populated macrostate has around 34% population, and there are 5 states larger than 5%. The top line is the amino acid sequence of hIAPP(11-25), with N-terminus is capped by acetyl group and C-terminus by amine group. The positively-charged residues (¹¹R and ¹⁸H) are in red color.



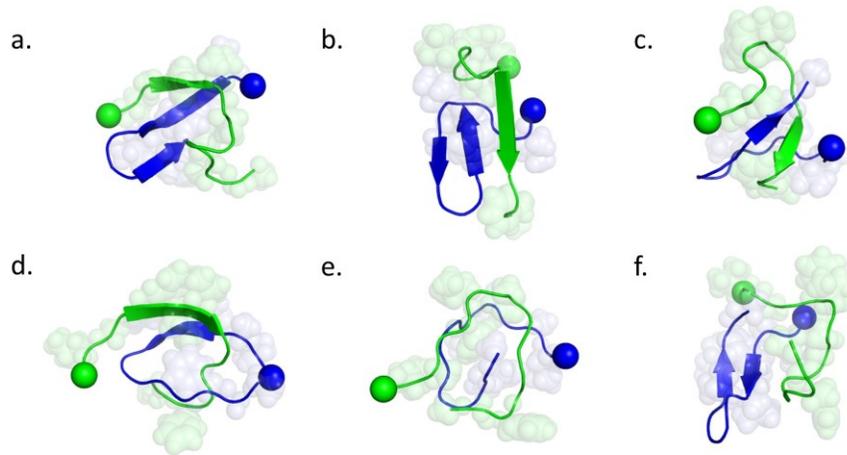
SI Fig. 4 The secondary structures (left column), contact map (middle column), and the representative structures (right column) of the 4th to 6th most populated macrostates. (a) The 4th most populated macrostate. One chain has a β -hairpin structure with β -bend- β secondary structures, and the two chains form contacts at the N-termini. (b) The 5th most populated macrostate has bend and turn structure and its inter-molecular contacts are around the C-termini. (c) The 6th most populated macrostate has some bend and β structures and forms antiparallel inter-molecular contacts in half of its sequence, while the rest parts separate in the opposite directions.



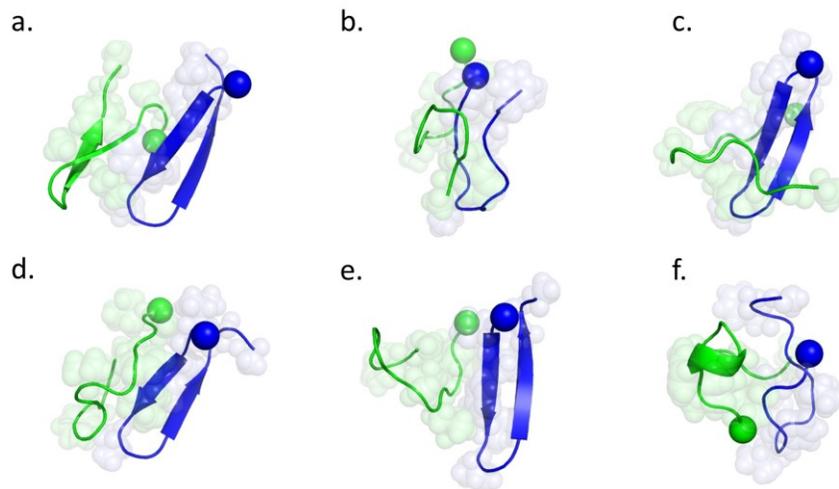
SI Fig. 5 The representative structures of the top most populated macrostate. (a-f) The top 6 most populated microstates in this macrostate. The structures are shown in cartoon representation and the hydrophobic residues are shown in transparent spheres. The N-termini of the two chains are shown in solid sphere. The Chain-1 is colored in blue and Chain-2 in green. The most populated macrostate has antiparallel extended cross- β structures with a small bend in the middle region.



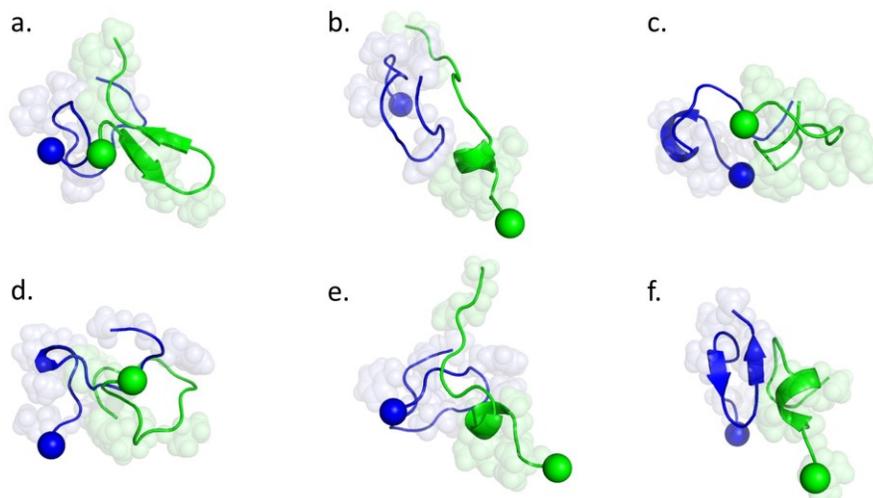
SI Fig. 6 The representative structures of the 2nd most populated macrostate. (a-f) The top 6 most populated microstates in this macrostate. The structure representation is the same as SI Fig. 4. The chains in the 2nd most populated macrostate form U-loop and their termini are in contact.



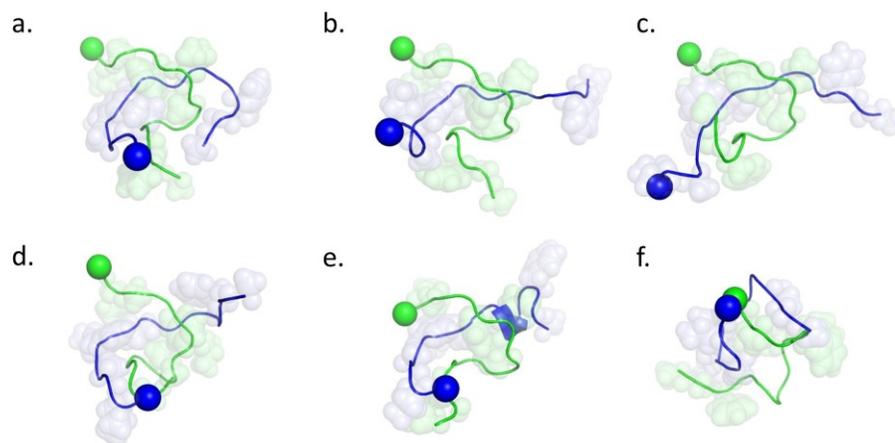
SI Fig. 7 The representative structures of the 3rd top most populated macrostate. (a-f) The top 6 most populated microstates in this macrostate. The structure representation is the same as SI Fig. 4. The chains in the 3rd most populated macrostate form U-loop and cross into each other with their middle regions far away.



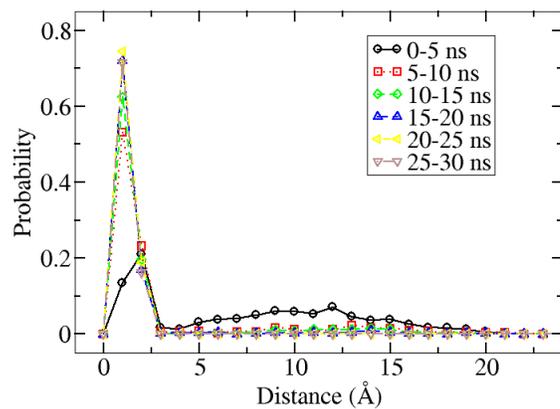
SI Fig. 8 The representative structures of the 4th most populated macrostate. (a-f) The top 6 most populated microstates in this macrostate. The structure representation is the same as SI Fig. 4. In this macrostate, one chain (blue) forms β -hairpin structure and the other (green) is flexible. The two chains form contacts at the N-termini.



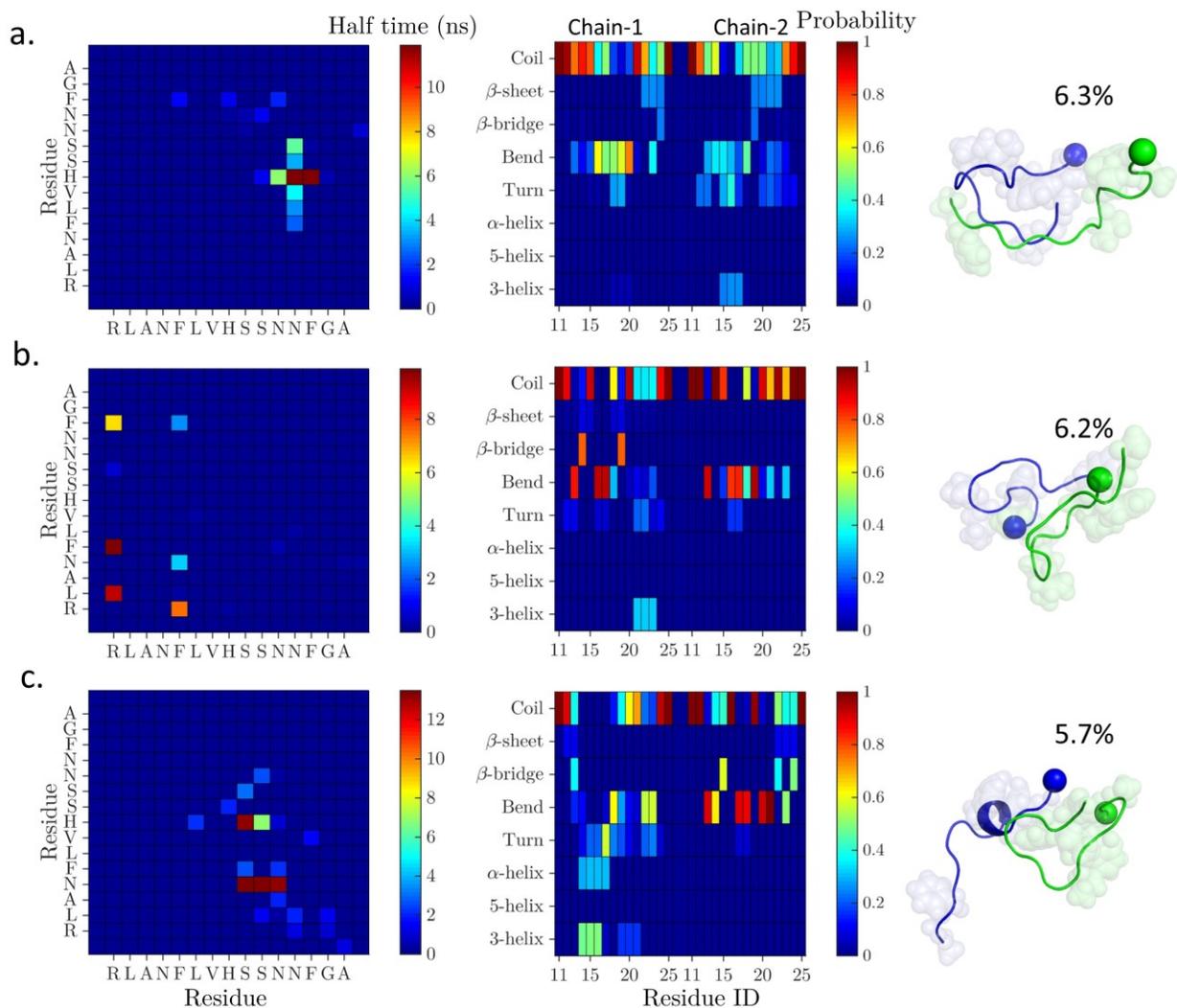
SI Fig. 9 The representative structures of the 5th most populated macrostate. (a-f) The top 6 most populated microstates in this macrostate. The structure representation is the same as SI Fig. 4.



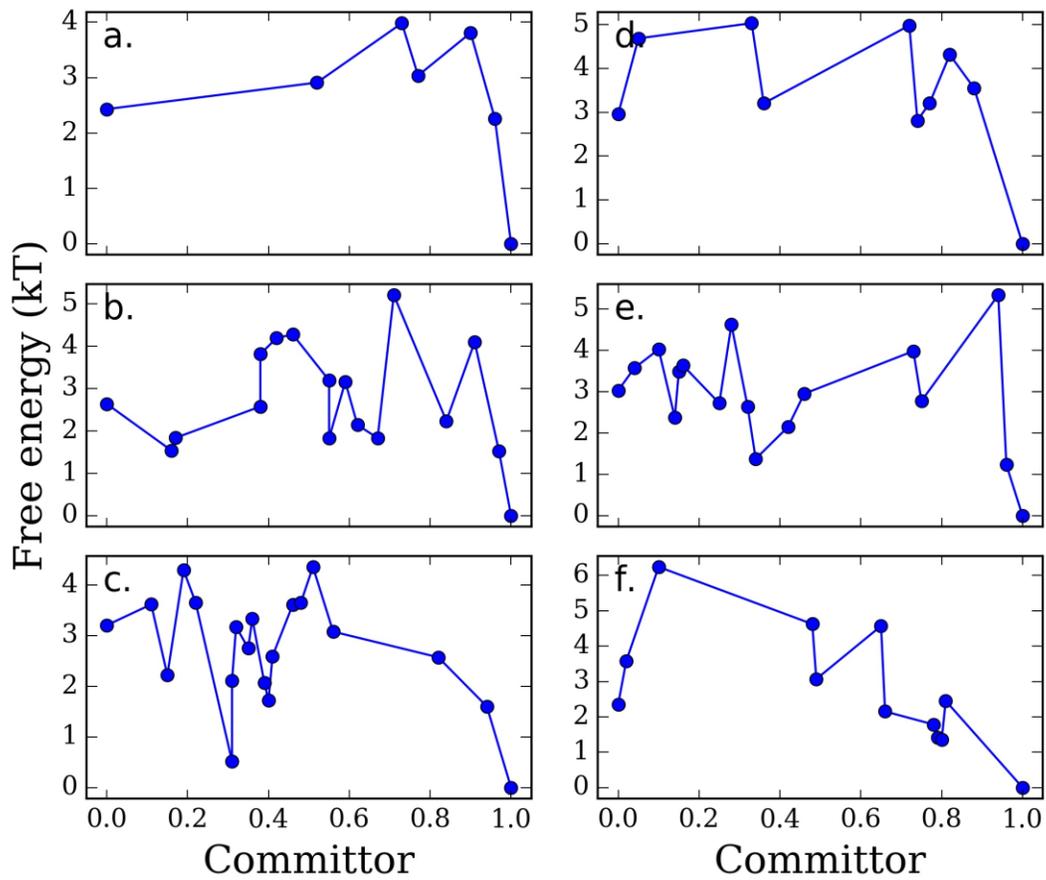
SI Fig. 10 The representative structures of the 6th most populated macrostate. (a-f) The top 6 most populated microstates in this macrostate. The structure representation is the same as SI Fig. 4. Each chain has a U-shape loop conformation and has the antiparallel inter-molecular arrangement in half of its sequence, while the other half goes apart in opposite directions.



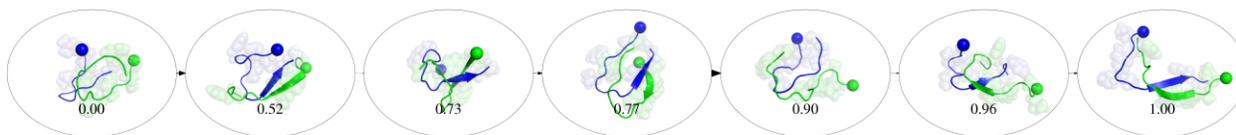
SI Fig. 11 The histogram of the minimum distance between two chains different time window from the separate (center of mass distance larger than 30 Å) conformations. Each curve represents a time window. From 5 to 10 ns, there exist little separate conformations, as shown in the red curve.



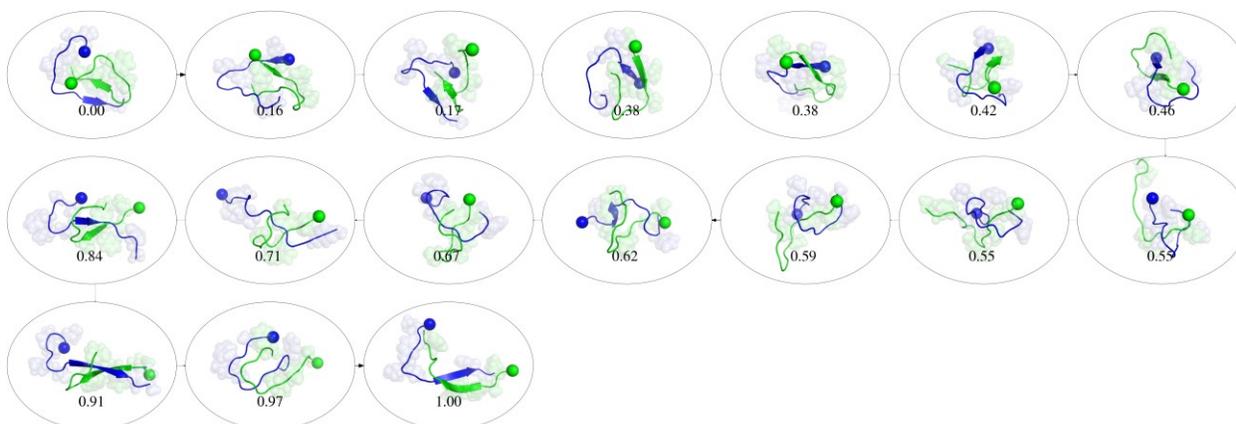
SI Fig. 12 The initial collapse process for the 4th to 6th most probable pathway. (a) The 4th most probable initial collapse process. (b) the 5th most probable (c) the 6th most probable. The left column shows the life time of the inter-molecular residue contacts in the initial collapse process. The middle column shows the secondary structures of the dimer conformations in the initial collapse process. The right column shows the representative conformations for each of the initial collapse processes.



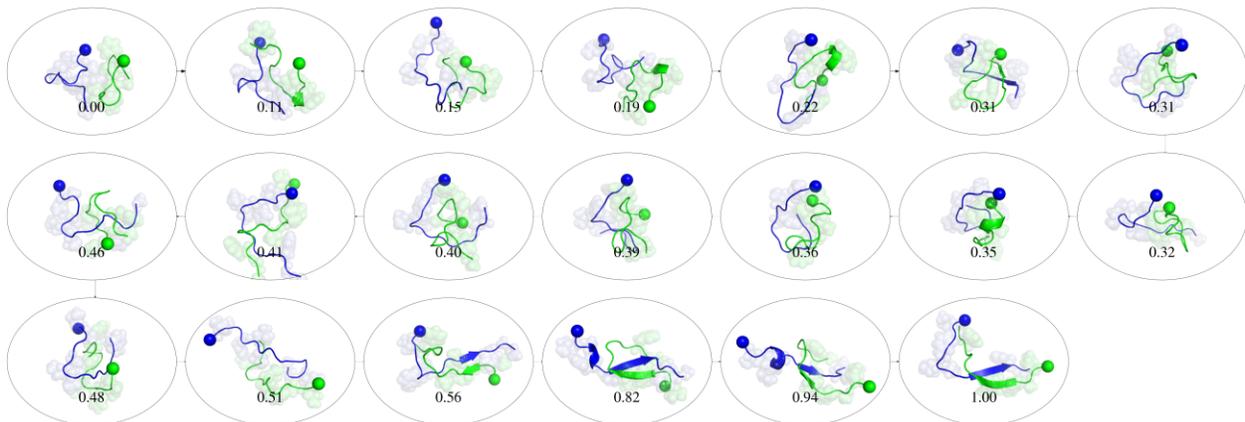
SI Fig. 13 The microstate free energy along the top 6 most probable pathways, with the free energy of the final state as the reference. (a-c) the top 3 pathways respectively (d-f) the top 4th to 6th pathways respectively.



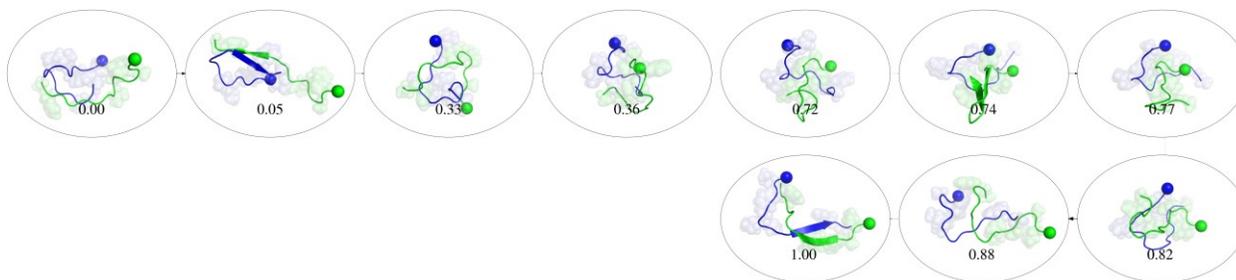
SI Fig. 14 Detailed structures of the top most probable pathway. The representation of structures is the same as SI Fig. 4(a). The arrow widths are proportional to the net-flux of the corresponding transitions. The labels on the bottom indicate the forward committer probability of each state.



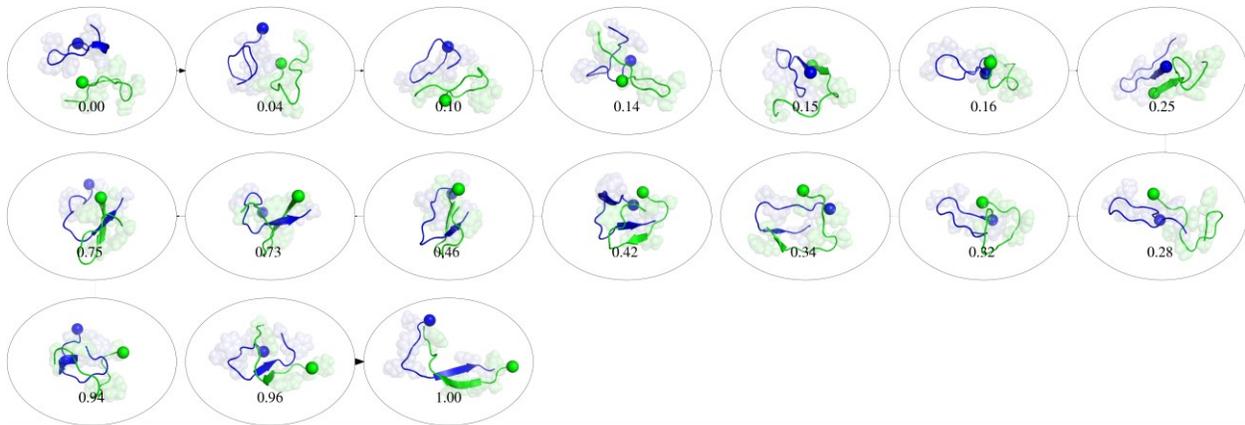
SI Fig. 15 Detailed structures of the 2nd most probable pathway. The representation of structures is the same as SI Fig. 4(a). The arrow width is proportional to the net-flux of the transition. The states are ordered in snake-shape, and the labels on the bottom indicate the forward committor probability of each state.



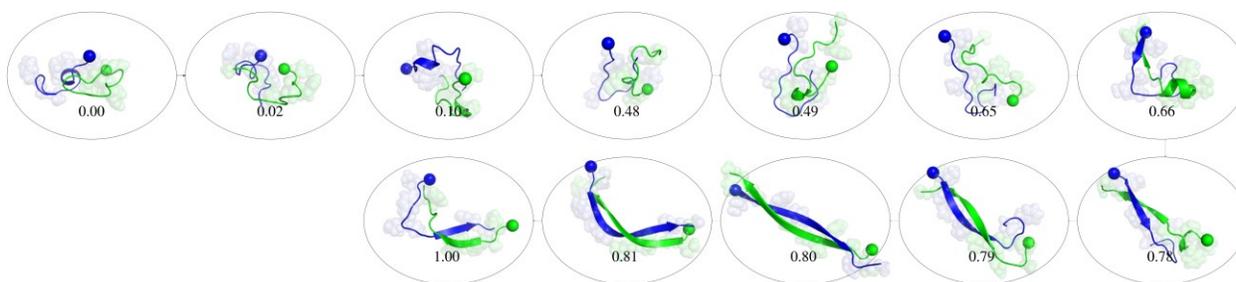
SI Fig. 16 Detailed structures of the 3rd most probable pathway. The representation of structures is the same as SI Fig. 4(a). The arrow width is proportional to the net-flux of the transition. The states are ordered in snake-shape, and the labels on the bottom indicate the forward committor probability of each state.



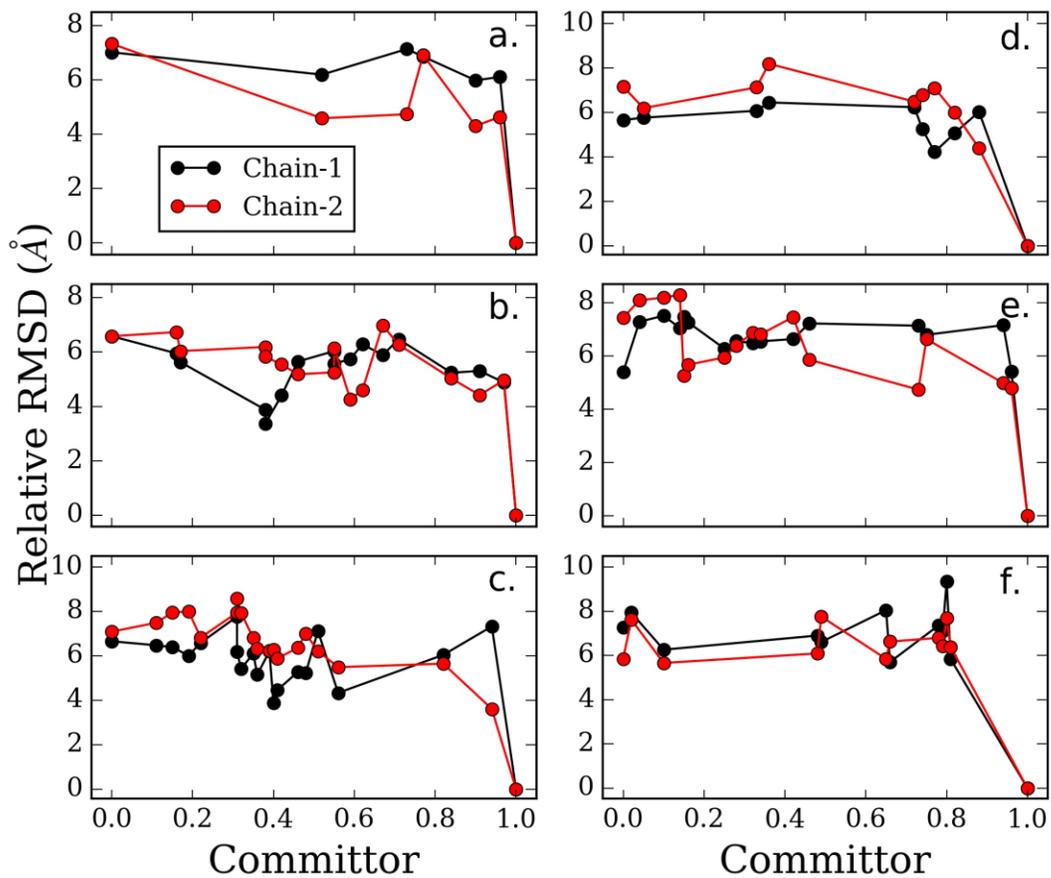
SI Fig. 17 Detailed structures of the 4th most probable pathway. The representation of structures is the same as SI Fig. 4(a). The arrow width is proportional to the net-flux of the transition. The labels on the bottom indicate the forward committer probability of each state.



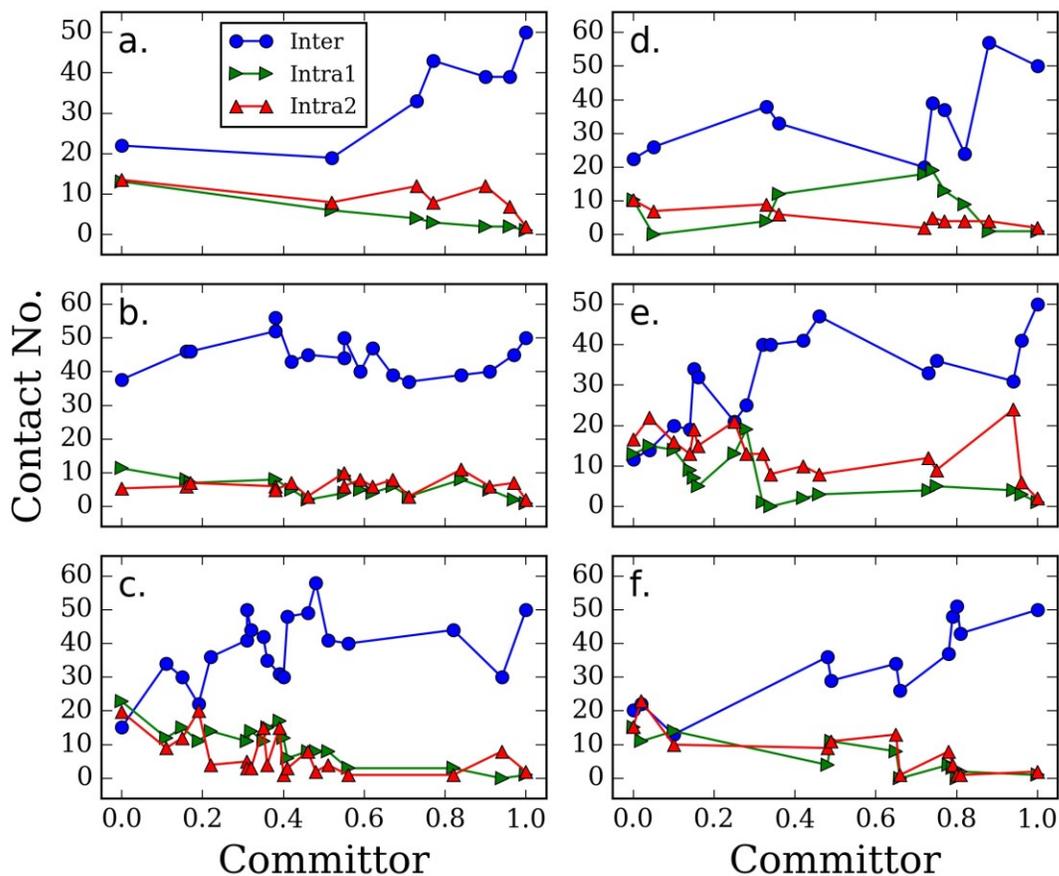
SI Fig. 18 Detailed structures of the 5th most probable pathway. The representation of structures is the same as SI Fig. 4(a). The arrow width is proportional to the netflux of the transition. The states are ordered in snake-shape, and the labels on the bottom indicate the forward committor probability of each state.



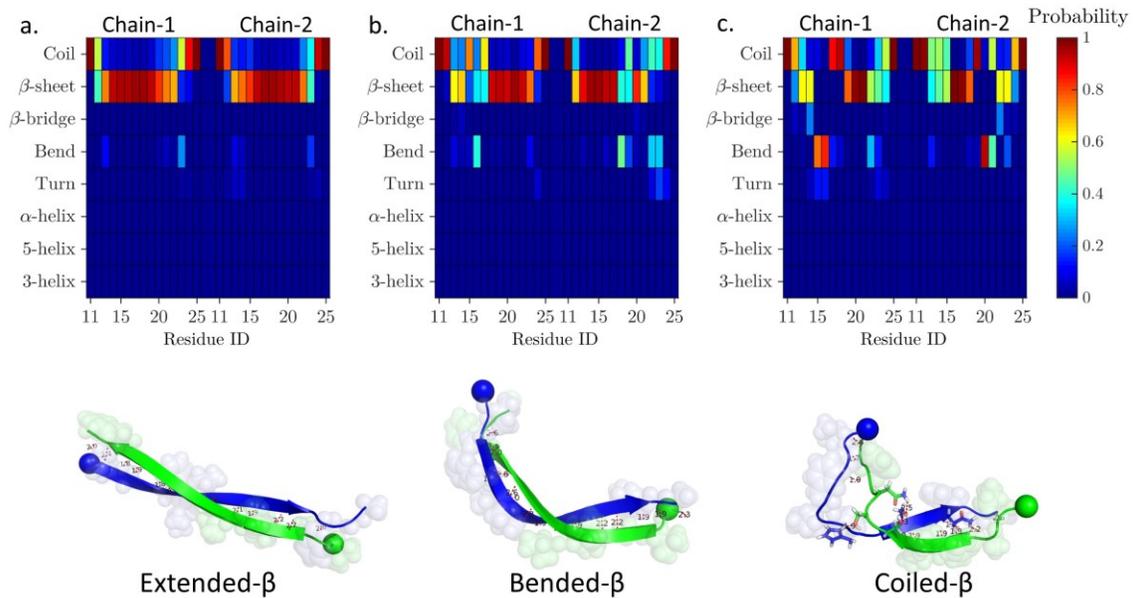
SI Fig. 19 Detailed structures of the 6th most probable pathway. The representation of structures is the same as SI Fig. 4(a). The arrow width is proportional to the netflux of the transition. The labels on the bottom indicate the forward committor probability of each state.



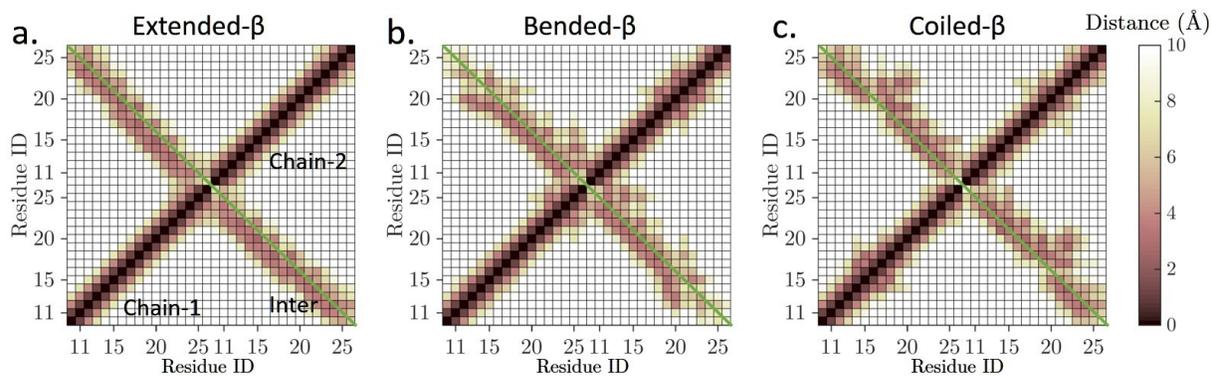
SI Fig. 20 The change of the RMSD of each chain between the states along the top 6 pathways and the final cross- β antiparallel structure. The black curves are for the Chain-1 and the red ones are for the Chain-2.



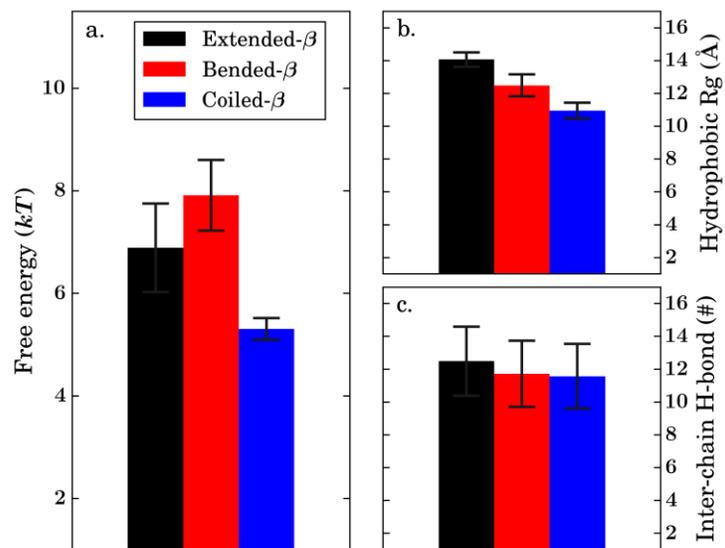
SI Fig. 21 The change of the residue-residue contact numbers along the top 6 pathways. The inter-molecular contact numbers are plotted in blue circles, the intra-molecular ones of the Chain-1 plotted in green triangles, and those of the Chain-2 plotted in red triangles. Along each pathway, the inter-molecular contact number increases with fluctuation, while the intra-molecular one decreases.



SI Fig. 22 Comparison of the secondary structures of the three antiparallel β -sheet states (a) the fully-extended- β state, labelled as ‘Extended- β ’ (b) the ‘Bended- β ’ intermediate state (c) the ‘Coiled- β ’ final state. The top row shows the residue-based secondary structures of each state, while the bottom one shows their representative conformations. The conformations are shown in cartoon representation and the sidechains of hydrophobic residues are shown in transparent spheres. The N-termini of the two chains are shown in solid sphere. Chain-1 is colored in blue and Chain-2 in green. The inter-molecular hydrogen bonds are shown in small red sticks, on which their bond lengths are labelled. The sidechains participating in the inter-molecular hydrogen bonds are shown in sticks representation in (c).



SI Fig. 23 Comparison of the residue-based contact maps of the three antiparallel β -sheet states. (a) the ‘Extended- β ’ state (b) the ‘Bended- β ’ state (c) the ‘Coiled- β ’ state. The anti-diagonal green lines indicate the perfect all-to-all inter-chain antiparallel association.



SI Fig. 24 Comparison of the thermo-stability of the three antiparallel β -sheet states. The black bars represent for the ‘Extended- β ’ state, the red ones for the ‘Bended- β ’ state, and the blue ones for the ‘Coiled- β ’ state. (a) the free energy (b) the Rg of the hydrophobic residues (c) the number of inter-chain hydrogen-bonds.