Multistep molecular mechanisms of Aβ16-22 fibril formation revealed by lattice Monte Carlo simulations

Phuong H Nguyen, Philippe Derreumaux

Accepted to J. Chem. Phys. 10.1063/5.0149419

Multistep Molecular Mechanisms of Aβ16-22 Fibril Formation Revealed by Lattice Monte Carlo Simulations

Phuong H. Nguyen,[1] Philippe Derreumaux [1,2*]

1 CNRS, Université Paris Cité, UPR 9080, Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild, 13 rue Pierre et Marie Curie, 75005 Paris, France.

2 Institut Universitaire de France (IUF), 75005, Paris, France.

Corresponding author: philippe.derreumaux@ibpc.fr

Abstract: As a model of self-assembly from disordered monomers to fibrils, the amyloid-β fragment Aβ16-22 was subject to past numerous experimental and computational studies. Because dynamics information between milliseconds and seconds cannot be assessed by both studies, we lack a full understanding of its oligomerization. Lattice simulations are particularly well suited to capture pathways to fibrils. In this study, we explored the aggregation of 10 Aβ16-22 peptides using 65 lattice Monte Carlo simulations, each simulation consisting of three billion steps. Based on a total of 24 and 41 simulations which converge and do not converge to the fibril state, respectively we are able to reveal the diversity of the pathways leading to fibril structure and the conformational traps slowing down fibril formation.

Keywords: amyloid, aggregation, simulations, lattice model, Alzheimer's disease

I.INTRODUCTION

A hallmark of amyloid proteins of different amino acid composition and length, such as Aβ40/Aβ42 and tau linked to Alzheimer's disease, alpha-synuclein linked to Parkinson's disease, and human islet amyloid polypeptide (hiAPP) linked to type II diabetes is the aggregation into amyloid fibrils with a cross-β structure and intermolecular H-bonds parallel to the fibril axis.[1] A large body of studies suggested that the early formed oligomers are the main causes of these neurogenerative diseases,[2,3] and amyloid fibrils acted as a reservoir of soluble oligomers.[4]

The lack of stable secondary and tertiary structure of monomers and small oligomers prevented high-resolution structure determination.[5] In contrast the structures of Aβ40/42 fibrils based on cryo-EM (electron microscopy) and ss-NMR (solid-state nuclear magnetic resonance) experiments revealed polymorphs with different intra- and intermolecular structures, suggesting Alzheimer's disease modifying strategies.[6,7]

Many studies based on Thioflavin T (ThT) fluorescence and size exclusion chromatography have emphasized the role of the residues 41 and 42,[8,9] residues 23-28,[10] and N-terminal truncations[11] on Aβ aggregation kinetics. Experimental studies also showed that the central hydrophobic core (CHC) spanning residues 17-21 was pivotal for aggregation of the full-length Aβ peptides.[12,13] The Aβ16-22 peptide of sequence KVLFFAE forms a cross-β fibril structure with antiparallel β-strands and antiparallel β-sheets as determined by ss-NMR[14] and microcrystal X-ray diffraction,[15] which represents an ideal and simple model for experimental and computational techniques.

Electron spray ionization ion-mobility spectrometry mass-spectrometry (ESI-IMS-MS),[16] and transmission electron microscopy (TEM) fibrillation experiments on Aβ16-22 blocked by acetyl and amine groups at concentrations from 10 to 200 μM at 277–330 K[17] suggested a primary one-single step nucleation (1SN) mechanism, i.e., fibril formation directly from the assembly of early ordered oligomers.[18,19] Under the conditions used in this study, Aβ16-22 formed fibrils within 5 min.[16] The 1SN mechanism was confirmed by ThT fluorescence on the N- and C-terminal unprotected Aβ16-22 peptide at neutral pH and a concentration of 30 μM and above under various conditions (salt and 2-propanol) showing fibrillation within seconds.[20] The 1SN mechanism was also corroborated by the coarse-grained PRIME model with implicit solvent coupled to discontinuous molecular dynamics (DMD)[16] and the PACE-ASM (atomistic representation of the peptide with Martini solvent) coupled to molecular

dynamics (MD) simulation.[21] The PRIME-DMD simulation formed amyloids fibrils after 12 μs at 326 K and a concentration of 10 mM. It has to be noted that the DMD approach neglects hydrodynamics effects which impact the fluctuations and life-time of the oligomers.[23-25] The PACE-ASM simulation converged even faster to fibrils on the order of 3 μs at 330 K.

Modifications of the experimental conditions alter Aβ16-22 oligomerization mechanisms. While the previous ESI-IMS-MS study used 100 mM ammonium acetate and a peptide concentration of 40 μM,[16] Lynn et al. by using 40% acetonitrile in water, a concentration of 1.3mM, circular dichroism (CD) and Fourier transformed infrared (FTIR) spectroscopy found out-of-register antiparallel β-sheets before formation of the full in-register of the strands.[26] Atomic force microscopy (AFM) – IR spectroscopy on 1mM unblocked Aβ16-22 peptide in 10mM phosphate buffer or 10mM HCl at 37°C proposed that the early aggregates displayed two β-sheets: a mixture of parallel antiparallel β-sheet and a fully antiparallel β-sheet.[27] It is interesting that tip-enhanced Raman spectroscopy on Aβ42 showed a late-stage transition from antiparallel β-sheet in protofibrils to parallel β-sheet in fibrils.[28]

Alternatively, the two-step nucleation 2SN (hydrophobic collapse leading to formation of amorphous aggregates followed by reorganization to β-sheet aggregates and to fibrils) mechanism was observed on unblocked Aβ16-22 peptide at concentrations of 0.6 mM/4 mM using fluorescence imaging.[29] Several theoretical works showed that 1SN and 2SN mechanisms depend on temperature, concentration,[30] and a small change in the intermolecular interactions,[18] and both 1SN and 2SN mechanisms can occur simultaneously.[30] Monte Carlo (MC) and path sampling simulations based on coarse-grained models with implicit solvent also proposed the 2SN mechanism.[19,31]

Atomistic MD simulations were used to determine the stability of preformed single layer of β-sheet and double-layer β-sheets with different numbers of β-strands of Aβ16-22 in each layer.[32-34] Single layers were found unstable, and rotation of one sheet with respect to the other by 90° was detected due to the intrinsic flexibly of each peptide and the dynamics of water.[33,34] MD simulations were also employed to understand Aβ16-22 fibril growth and provided evidence of the dock and lock mechanism for fibril elongation.[35] Molecular simulations on large oligomers showed that Aβ16-22 formed metastable well-structured β-barrels and perpendicular β-sheets.[36-38]

Overall, the experimental times for primary nucleation and secondary nucleation composed of fragmentation and surface-catalysed effects[8] are out-of-reach by standard

atomistic MD simulations in explicit solvent. Enhanced sampling methods such as replica exchange molecular dynamics (REMD),[39,40] replica exchange with solute tempering,[41,42] metadynamics,[43,44] and path sampling[31] with different protein representations and force fields were used to overcome the free energy barriers involved in each microscopic event. However, metadynamics efficiency depends on the definition of collective variables, path sampling technique depends on the order parameters, and the number of replicas used in REMD increases substantially with the number of degrees of freedom.

Within the context of aggregation simulation studies, shape-free theory for the self-assembly kinetics in macromolecular systems was also developed,[45] and lattice MC simulations were often used to probe fibril formation.[46-51] Past MC simulations showed that the energy difference between the monomeric native and fibril-prone states was the main determinant of fibril formation.[51] We recently added the OPEP force field[52] to the Abeln's lattice MC model and program.[53] After optimizing the energy parameters to Aβ16-22 dimer and verifying their robustness on Aβ16-22 trimer, Aβ37-42 dimer and trimer by reproducing the energy landscapes obtained by all-atom REMD simulations in explicit solvent, we constructed the free energy landscape (FEL) of 10 Aβ16-22 peptides projected on the nematic order parameter P2 (see Eq. 3 in Ref. 53 for its precise implementation using a = 0.3 nm the distance between two consecutive Cα atoms) and the number of intermolecular H-bonds using replica exchange Monte Carlo (REMC) simulations. We showed that the FEL of Aβ16-22 peptides was dominated by aggregates matching the experimental fibril structure characterized by two antiparallel sheets with antiparallel strands.[54] It has to be noted that the predicted FELs of 10, 20 and 40 Aβ37-42 peptides were totally different, as Aβ37-42 fibrils form antiparallel β-sheets and parallel β-strands.[54,55] In this study, we investigate the aggregation mechanisms of 10 Aβ16-22 peptides using 65 lattice OPEP MC simulations.

II. MATERIAL AND METHODS

We used the cubic lattice model and program developed by Abeln et al. which include two particles per residue (Cα, and side chain bead).[53] Each residue is located on the vertex of a cubic lattice, each central residue has four possible side chain directions, and each terminus has five side chain directions. A H-bond is formed if two residues are in contact and the side chains are oriented in the same direction. The side chains interact when they are directed toward each other or when they point in the same direction in a parallel fashion. The Abeln's

energy used in Ref. 53 consists of four terms: two-body H-bond interactions, pairwise interactions between the amino acid side chains, a steric term to prevent clash between consecutive side chains oriented in the same direction, and a solvent term (E-solvent) when a residue touches at least one empty vertex.[53] The reader can refer to Ref. 53 for the simulation degrees of freedom (position- and side-chain vectors) and the formulation of the potential energy terms.

As in our previous REMC simulation, we replaced the Abeln's pairwise residue interactions by the OPEP pairwise residue interactions, added the four-body H-bond OPEP term, and set E-solvent to zero.[54] The reader can refer to Ref. 54 for the implementation of the four-body H-bond term and the value of each parameter. In this study, we also used the N- and C-terminal unprotected Aβ16-22 peptide and a cubic box of 51 * 51* 51 lattice points leading to a concentration, *c*, of about 4.5 mM using the formula c =(N/NA*V) where N is the number of peptides, NA is the Avogadro's number and V is the volume in liter using a lattice spacing of 0.3 nm.

Starting from fully dispersed monomers, a total of 65 MC simulations was performed in the NVT ensemble with periodic boundary conditions for 2.8 10^9 steps with trial steps accepted according to the Metropolis criterion[56] at T* = 0.27. This temperature was selected because it is just below the temperature at which the heat capacity profile of 10-mers displays a peak using REMC simulations with 120 replicas.[54] Trial moves, which exactly match those in Abeln's work, include either internal moves changing the conformation of a backbone peptide (corner flip, end move, crank shaft, point (branch) rotation)), or rigid body moves, changing the position of the peptide to other objects (rotation, translation).[53] Local moves to change the states between strand and coil and change the side chain directions are also performed. Overall, at each MC step, a single local trial move and a global trial move (including point rotations) with the probability P_{global} = 0.1 are performed, see Ref. 53 for more details. Cluster moves where several peptides are updated simultaneously are not attempted, while moves of backbone and side chain particles are updated simultaneously.

We emphasize that this ensemble of update moves and attempted probabilities at each MC iteration in Abeln's lattice model program captured the folding of a designed protein of 35 residues with high β-sheet content, and amyloid formation of ten TFTFTFT peptides, albeit no aggregation pathways were described.[53] Our past MC simulations with Abeln's moves were able to describe accurately the equilibrium states of Aβ16-22 and Aβ37-42

amyloid fibrils with distinct β-strand orientations.[55] Multiple association and dissociation of the chains from transiently formed oligomers were also observed. All together the ensemble of MC moves provides evidence, but does not prove, the very good sampling efficiency of the NVT Monte Carlo simulations

Conformational transitions along MC trajectories were monitored by the P2 parameter, the number of H-bonds (# Hbond), the total energy in units of k_BT, the size of the β-sheets, the content of mixed antiparallel/parallel or fully antiparallel strands in all β-sheets and in particular within one layer of five-stranded β-strands using the cosine of the angles between vectors, and the probability distributions P(total cosine angle, β-sheet size). The β-sheet size n is defined as the number of n β-strands connected consecutively by at least five H-bonds.[54] The total cosine of the angle is the sum of all cosine angles between the end-to-end vectors of two consecutive hydrogen bonded strands.[54] Inspection of all trajectories allowed us to identify the number of converging and non-converging simulations, the former (latter) refers to runs that contain (do not contain) a native fibril at the end of the simulation.

III. RESULTS AND DISCUSSION

Detailed Insights into 4 simulations. We first present an energetic and a conformational description of the aggregates in runs R1 and R3 which converge to the fibril state, and runs R2 and R4 which do not converge to the fibril state within 2.8 10^9 MC steps. Although the description of the first four individual MC trajectories may not reflect the total 65 simulations, they help us identify a list of progress variables to distinguish the simulations. The detailed descriptions of few individual MC trajectories do not imply that the trajectories resemble actual physical processes. Their purposes are to provide the reader with a better understanding what happens in the MC simulations.

In run R1, during 2.06 10^9 MC steps, we see dispersed and disordered monomers (states 1 and 2 in Figure 3, P2 on the order of 0.1 in Figure 1) which evolve at 2.08 10^9 MC steps to an antiparallel 3-stranded β-sheet, a two-stranded antiparallel β-sheet and five disordered monomers (state 3 in Figure 3; P2 =0.53 and # Hbond = 5 in Figure 1). These two- and three-stranded β-sheets coalesce at 2.16 10^9 MC steps to generate a cross-β structure made of seven and three chains (state 4 in Figure 3; P2 = 1, E = -81.2 k_BT and # Hbond = 23 in Figure 1), indicating the role played by the surface-catalysed secondary nucleation mechanism. Next, three chains at the extremity of the seven-stranded β-sheet detach at 2.26

10^9 MC steps by a fragmentation event and convert to two unfolded chains and one chain packed correctly to the second three-stranded β-sheet (state 5 in Figure 3; P2 = 0.7 and # Hbond = 26 in Figure 1). It has to be noted that state 5 with E = -66.3 k_BT has a much higher energy than state 4, and this transition requires 1 10^8 MC events. This oligomer finds a pathway of decreasing energy to form at 2.81 10^9 MC steps the perfect fibril state (P2 = 1, # Hbond = 37 and cosine angle = -8 in Figure 1, state 6 in Figure 3) with an energy of -87.4 k_BT. Overall, in the full R1 trajectory there is a small amount of mixed parallel-antiparallel β-sheet content in five stranded β-sheets with 0 < cosine angles < 4 (Figure 1).

The initial event in run R2 involves at 0.25 10^9 MC steps a 3-stranded parallel β-sheet and seven free monomers (state 1 in Figure 4; E = -19.2 k_BT and P2= 0.33 in Figure 1) which convert at 0.48 10^9 MC steps to one single β-sheet made of nine out-of-register mixed antiparallel/parallel β-strands (state 2 in Figure 4, E = -59.4 k_BT, P2= 0.85, and # Hbond = 38 in Figure 1). From there and the rest of the simulation up to 2.81 10^9 MC steps (states from 3 to 6 in Figure 4), the system remains trapped into a non-native cross-β structure, with the two sheets antiparallel as in the native state, but incorrect orientation of the strands in the two sheets, P2 varying from 0.69 to 1, # Hbond varying between 23 and 34 and, and reaching a final energy of -83.0 k_BT (Figure 1). Overall, in the full R2 trajectory, the majority of the cosine of all angles varies between 2 and -2, indicating a very high percentage of mixed parallel antiparallel content in five stranded β-sheets (Figure 1).

The initial step in run R3 at 0.86 10^8 MC steps involves the formation of a six-stranded antiparallel out-of-register β-sheet (state 1 in Figure 5; E = -47.0 k_BT, P2= 0.58, # Hbond = 25 and cosine =-8 in Figure 2). This is followed at 4.1 10^8 MC steps by the formation of a single β-sheet made of ten antiparallel out-of-register peptides (state 2 in Figure 5; E = -66.5 k_BT, P2= 0.93 and # Hbond = 39 in Figure 2). Next the β-sheet fragments at 6.8 10^8 MC (state 3 in Figure 5, with E = -39.0 k_BT), and the oligomer associates again into an 8-stranded antiparallel β-sheet at 1.15 10^9 MC steps (state 4 in Figure 5, with E = -60.9 k_BT). Fragmentation and dissociation/ association are repeated, see state 5 (E = -50.1 k_BT at 1.45 10^9 MC steps) and state 6 (E = -46.0 k_BT at 1.66 10^9 MC steps in Figure 3). Then at 1.74 10^9 MC steps the system is able to form two antiparallel β sheets consisting of four and five antiparallel β-strands (state 7 in Figure 3 with E = -76.0 k_BT and # H-bond = 39). It takes about 1.1 10^9 MC step for the system to add the last monomer and reaches the native fibril state (state 8 in Figure 5 at 2.8 10^8 MC) with an energy

of -88.6 k_BT (Figure 2). Similarly, to run R1, the full R3 trajectory displays a few mixed parallel-antiparallel content in five stranded β-sheets (Figure 2).

Run R4 starts by the formation of a mixed parallel antiparallel seven-stranded β-sheet at 3.0 10^8 MC steps (state 2 in Figure 6, E = - 49.3 k_BT) followed by an in-register antiparallel ten-stranded β-sheet (state 3 in Figure 6, E = - 78.3 k_BT, cosine = -8) at 6.1 10^8 MC steps. Then, the oligomer explores higher energy states (state 4 in Figure 6, with E = - 41.1 k_BT in Figure 2), characterized by an organized tetramer, and an organized dimer that evolve to a lower energy state (state 5 with E = - 76.0 k_BT). This is followed by high fluctuations in energy, P2 and # Hbonds allowing the formation of two in-register antiparallel eight-stranded and nine stranded β-sheets (states 6 and 7 in Figure 6 at 1.4 and 1.9 10^9 MC steps with E = - 60 and -67.1 k_BT in Figure 2, respectively) which depolymerizes by one monomer at 2.8 10^9 MC steps with E = -57.9 k_BT (state 8). Overall, run R4 shows a very low content of parallel, antiparallel or mixed strand content in five stranded β-sheets (Figures 2 and 6).

Several Conformational Traps Slow Down Fibril Formation. Among 65 simulations, 41 simulations did not converge, and their final structures can be grouped into seven clusters. The structures of clusters C2 to C7 are shown in Figure 7. Cluster C1 observed in 6 runs displays an oligomer size distribution peaked at free monomers (population of 80*%) with smaller peaks for 2-, 3-, 4- and 5-mers, representing a population of 20%. Cluster C2 observed in 13 runs displays N-stranded β-sheets with 10-N dispersed peptides, N taking values of 6, 7, 8, 9 and 10. The C3 and C4 clusters, observed in 11 and 2 simulations, respectively consist of two five-stranded β-sheets with mixed antiparallel parallel orientations, either in the middle of two sheets (inner peptides, cluster C3) or in a full β-sheet (outer and inner peptides, cluster C4). It is noted that the frequency of incorrectly orientated inner peptides is higher that the frequency of misoriented outer peptides. Cluster C5 observed in four runs consists of out-of-register antiparallel β-sheets. Cluster C6 observed in two runs consists of out-of-register mixed antiparallel-parallel β-sheets. It is to be noted that cluster C5 also displays three β-strands with two corner flips. Finally, cluster C7 observed in three runs shows amorphous aggregates with unfolded peptides.

Comparison with experiment and past simulations. The seven clusters explain the long lag phase observed by ThT fluorescence curves.[8] Clusters C1 and C7 display a majority of the peptides remaining dispersed, forming transient small oligomers (C1) or showing aggregates with unfolded peptides (C7). The slow decrease in monomer concentration as a

function of time was discussed experimentally for many amyloid peptides.[3,9,70] Previous atomic and coarse-grained simulations never reported the existence of cluster C1 in long trajectories at mM concentration. Cluster C7 is consistent with the low β-signal during the lag phase observed by means of experiments[9] and MD and REST2 simulations.[42]

We never trapped stable oligomers of small sizes with β-sheet (cluster C2) experimentally. Cluster C2 is, however, consistent with the 1SN mechanism,[18-20] and past lattice MC simulations which identified that the free energy barrier that a nascent fibril has to overcome is associated with change in width.[57] Using cuboidal building unit and MC simulations, it was also evidenced that nucleation is linked to formation of multiple β-sheets.[58]

There is strong similarity between the clusters C3-C4 and both experimental and past computational results. Clusters C3 and C4 involving mixed antiparallel parallel strands are supported by the fact that the β-sheet competition between parallel and antiparallel strands is already encoded in small oligomers of Aβ16-22 computationally.[59-61] Mixed orientated strands were further described by AFM-IR spectroscopy,[27] MD, MC and bias-exchange metadynamics simulations using both atomistic and coarse-grained representations and implicit/explicit solvent of many amyloidogenic peptides.[44,62-70] Furthermore, the Aβ16-22 variant with the C-terminal residue E mutated by residue Q forms formed parallel β-strands by ss-NMR,[71] indicating a small free energy difference between the two types of strand patterns, and simulations revealed that parallel strands become dominant only in the large oligomers.[72] In addition, a novel mixed antiparallel-parallel β-structure formed by Aβ40 on GM1 clusters was assessed by ss-NMR,[73] and NMR data also suggested Aβ preglobulomer and globulomer with a mixed parallel and antiparallel structure.[74]

Finally, the clusters C5 and C4 have been described experimentally and computationally. Clusters C5 and C6 with out-of-register strand arrangements are supported by Aβ16-22 MD simulations and Markov state models of fibril growth which predicted that amyloid assembly is dominated by misregistered kinetic traps,[75] isotope-edited IR and FTIR spectroscopies,[26,76,77] and coarse-grained activation-relaxation simulations.[78,79] In addition, Aβ16-22 peptide shows different H-bond registers as a function of pH experimentally,[26] indicating that both the in-register and out-of-register strands are present in the full conformational spectrum of the oligomers. It is interesting that the crystal structure of hIAPP fragments revealed packing motifs of out-of-register β-sheets,[80] and out-of-register β-sheets have been proposed to be part of a toxic pathway.[81]

Probability Distributions (PD) and Dynamics of Oligomerization. There are significant differences between the two PDs using the full trajectories of both the 24 converging (Fig. 8A) and 41 non-converging simulations (Fig. 8B). The PDs represent the probability distributions P (β-sheet size, cosine angle) obtained from a histogram of the trajectories. Table 1 describes the structural characteristics of the dominant β-sheet sizes during the full trajectories.

The converging trajectories explore mainly five states, representing 72% of all conformations. The native fibril state (s4) with two antiparallel β-sheets and a cosine angle of -8 has a population of 0.47. The other four transient states, which have a (β-sheet size, cosine angle) values of (3, -2) for s1; (4, -3) for s2, (5, -4) for s3, and (9, -8) for s'5, are all characterized by N antiparallel β-strands with N=3, 4, 5, and 9 within one single layer β-sheet.

A total of 9 transient states of populations varying from 0.028 (s'7 state) to 0.161 (s'4 states) and therefore representing 61% of all conformations, is revealed during the non-converging simulations. The s'4 state of (β-sheet size, cosine angle) of (5, -4) corresponds to a native single β-sheet of five antiparallel β-strands, but this state is not sufficient to lead to a native fibril. Rather the oligomer evolves either to two β-sheets with mixed orientated β-strands (s'5 state, population of 0.052) or two parallel β-sheets (s'6 state, population of 0.053). It is also noted that the s'1 state of population 0.08 consists of two antiparallel β-strands, the s'2 and s'3 states with populations of 0.114 and associated with β-sheet sizes of 5 consist of mixed parallel/antiparallel β-strands – these two states are not explored during the converging simulations - and the s'7, s'8 and s'9 states with a population of 0.179 correspond to one single layer β-sheet made of 8 and 9, and 10 strands which cannot fragment. It is to be noted that the final structures of clusters C1 to C7 obtained by non-converging simulations do not correspond to localised peaks because their probabilities are very small (< 0.02) considering a cumulative number of 114 (41 x 2.8) billion MC steps, and all clusters C1 to C7 do not contain a unique structure but multiple distinct substates. Finally, a two-dimensional probability distribution cannot cover the full conformational space. As an example, cluster C5 resembles the s'3 state in terms of cosine angle and β-sheet size, but has two layers.

To further understand the sampling in converging and non-converging simulations, we reported on the MC step evolution of the β-sheet sizes ranging from a two-stranded (2S) β-sheet to a five-stranded (5S) β-sheet. In non-converging simulations, the fluctuations are very small for 2S, 3S and 4S and centered about 1.5 for 2S, and 1 for both 3S and 4S (Figure 9A-C). For the 5S state, the β-sheet size rapidly reaches 1.5 and then slightly fluctuates about this

value for the rest of the simulations (Figure 9D). In contrast, in the converging simulations, the fluctuations are very large for 2S, 3S and 4S, fluctuating between 0 and 1 for 2S and 3S, and even between 0 and 2 for 4S (Figure 9A-C). For the 5S state, the β-sheet size rapidly reaches 1.5 and then progressively reaches 2, the native state. (Figure 9D) Overall, the dynamics is characterized by long MC lifetimes and short MC Lifetimes of the transient states in non-converging and converging simulations, respectively. This behavior is also reflected by the MC evolution of the total cosine angle (Figure 10) where the value slowly varies between 0 and -2 in non-converging simulations, while the value progressively descends from 0 to -8 with larger fluctuations in converging simulations.

IV. CONCLUSIONS

Using 65 on-lattice OPEP MC simulations, each simulation consisting of 3 billion steps, we elucidated the pathways leading to Aβ16-22 fibril formation. The simulations demonstrated the multistep molecular mechanisms and the heterogeneity of the ensemble of metastable intermediates leading to fibril formation. It is interesting that both the 1SN and 2SN mechanisms were observed in converging simulations. The computational study provided evidence of the role played by the surface of one β-sheet (surface-catalysis), fibril fragmentation, fibril depolymerization by one chain, and fibril elongation by one chain, consistent with the elementary aggregation events to explain the variation of ThT fluorescence upon aggregation time.[8]

The simulations revealed that the assembly of two β-sheets with mixed strand orientations or out-of-register strands and the formation of N-stranded β-sheet protofibril with N varying between 6 and 10 lead to conformational traps and have very long MC lifetimes. This runs in contrast to the converging simulations where the transient oligomers have short MC lifetimes. The importance of kinetics traps and kinetic control of amyloid self-assembly was revealed by many simulations,[42,44,59-70] including atomistic simulations of Aβ16-22 dimer[61] and coarse-grained protein model simulations of fibril formation.[82] The combined results of our MC simulations indicate a multi funnel free energy landscape, but the converging simulations find aggregation pathways avoiding kinetic traps.

We emphasize that all intermediates cannot form using our lattice model. These include α-helix or α-sheet oligomers which may play a role in the aggregation of Aβ42,[83,84] and also the β-barrels and perpendicular β-sheets which were shown to easily transform to cross-

β structures.[37,38,85] The lattice model does not consider the difference of (ϕ, Ψ) backbone values between the parallel and antiparallel β-sheets. Our Aβ16-22 model system might be more representative of other amyloid/pathogenic peptides with antiparallel fibrils than Aβ40 and Aβ42 peptides which form parallel β-sheets. Finally, NVT Monte Carlo simulations have limitations to study kinetics, and we cannot quantify the number of MC steps with actual real time. The obtained conformations can be used, however, in a Markov model to obtain the kinetics.

Overall, all the transient conformations and conformational traps diversify the ensemble of amyloid assemblies in aqueous solution. This heterogeneous diversity is likely to increase the number of adsorption and insertion modes of Aβ peptides in the membrane and to impact the interactions with protein membranes,[3,86,87] rendering the design of drugs targeting Aβ species more complicated.[3,88-91]

AUTHOR INFORMATION

ORCID:

Derreumaux: 0000-0001-9110-5585

Nguyen: 0000-0003-1284-967X, Phuong.nguyen@ibpc.fr

Notes

The authors declare no competing financial interest, and no conflicts.

ACKNOWLEDGMENTS

This work was supported by the “Initiative d’Excellence” program from the French State, Grant “DYNAMO”, ANR-11-LABX-0011. We thank Sanne Abeln for useful discussions, and Geoffrey Letessier for computational maintenance at Laboratoire de Biochimie Théorique.

REFERENCES

1.C.M. Dobson, Nature, **426**, 884-90 (2003).

2. D.J. Selkoe, and J. Hardy, Mol Med*.*, **8**, 595-608 (2016).

3. P.H. Nguyen, A. Ramamoorthy, B.R. Sahoo, J. Zheng, P. Faller, J.E. Straub, L. Dominguez, J.E. Shea, N.V. Dokholyan, A. De Simone et al., Chem Rev., **121**, 2545-2647 (2021).

4. A. Bigi, R. Cascella, F. Chito, and C. Cecchi, Bioessays, **44**, e2200086 (2022).

5. P.H. Nguyen PH, and P. Derreumaux, Biophys Chem., **264**, 106421 (2020).

6. J.X. Lu, W. Qiang, W.M. Yau, C.D. Schwieters, S.C. Meredith, and R. Tycko, Cell, **154**, 1257-68 (2013).

7. M. Willem, and M. Fändrich, Science, **375**, 147-148 (2022).

8. G. Meisl, X. Yang, E. Hellstrand, B. Frohm, J.B. Kirkegaard, S.I. Cohen, C.M. Dobson, S. Linse, and T.P. Knowles, Proc Natl Acad Sci U S A., **111**, 9384-9389 (2014).

9. G. Bitan, M.D. Kirkitadze, A. Lomakin, S.S. Vollers, G.B. Benedek, and D.B. Teplow, Proc Natl Acad Sci U S A., **100**, 330-335 (2003).

10. K.L. Sciarretta, D.J. Gordon, A.T. Petkova, R. Tycko, and S.C. Meredith, Biochemistry, **44**, 6003-6014 (2005).

11. C. Dammers, M. Schwarten, A.K. Buell, and D. Willbold, Chem Sci., **8**, 4996-5004 (2017).

12. L.O. Tjernberg, J. Näslund, F. Lindqvist, J. Johansson, A.R. Karlström, J. Thyberg, L. Terenius, and C. Nordstedt, J Biol Chem., **271**, 8545-8548 (1996).

13. A. Doig, and P. Derreumaux, Curr Opin Struct Biol., **30**, 50-56 (2015).

14. J.J. Balbach, Y. Ishii, O.N. Antzutkin, R.D. Leapman, N.W. Rizzo, F. Dyda, J. Reed, and R. Tycko, Biochemistry, **39**, 13748-59 (2000).

15. M.R. Sawaya, S. Sambashivan, R. Nelson, M.I. Ivanova, S.A. Sievers, M.I. Apostol, M.J. Thompson, M. Balbirnie, J.J. Wiltzius, H.T. McFarlane et al. Nature, **447**, 453-457 (2007).

16. S.J. Bunce, Y. Wang, S.E. Radford, A.J. Wilson, and C.K. Hall, AIChE J., **67**, e17101 (2021).

17. Y. Wang, S.J. Bunce, S.E Radford, A.J. Wilson, S. Auer, and C.K Hall, Proc Natl Acad Sci U S A., **116**, 2091-2096 (2019).

18. A. Šarić A, Y.C. Chebaro, T.P. Knowles, and D. Frenkel, Proc Natl Acad Sci U S A., **111**, 17869-17874 (2014).

19. A.J. Dear, T.C.T Michaels, G. Meisl, D. Klenerman, S. Wu, S. Perrett, S. Linse, C.M. Dobson, and T.P.J Knowles, Proc Natl Acad Sci U S A., **117**, 12087-12094 (2020).

20. M. Yamazaki, K. Ikeda, T. Kameda, H. Nakao, and M. Nakano, J Phys Chem Lett., **13**, 6031-6036 (2022).

21. X. Cai, and W Han, J Chem Inf Model., **62**, 2744-2760 (2022).

23. F. Sterpone, P. Derreumaux, and S. Melchionna, J Chem Theory Comput., **11**, 1843-1853 (2015).

24 M. Chiricotto, S. Melchionna, P. Derreumaux, and F. Sterpone, J Chem Phys., **145**, 035102 (2016).

25. M. Chiricotto, S. Melchionna, P. Derreumaux, and F. Sterpone, J Phys Chem Lett., **10**, 1594-1599 (2019).

26. M.C. Hsieh, C. Liang, A.K. Mehta, and D.G. Lynn, and M.A. Grover, J Am Chem Soc., **139**, 17007-17010 (2017).

27. S. Banerjee, D. Baghel, Md Hasan Ul Iqbal, and A. Ghosh, J Phys Chem Lett., **13**, 10522-10526 (2022).

28. E. Lipiec, J. Kaderli, J. Kobierski, R. Riek, K. Skirlińska-Nosek, K. Sofińska, M. Szymoński, and R. Zenobi, Angew Chem Int Ed Engl., **60**, 4545-4550 (2021).

29. Y. Liang, D.G. Lynn, and K.M. Berland, J Am Chem Soc., **132**, 6306-6308 (2010).

30. S. Auer, P. Ricchiuto, and D. Kashchiev, J Mol Biol., **422**, 723-730 (2012).

31. J.A. Luiken, and P.G. Bolhuis, J Phys Chem B., **119**, 12568-12579 (2015).

32. B. Ma, and R. Nussinov, Proc Natl Acad Sci U S A., **99**, 14126-14131 (2002).

33. U.F. Röhrig, A. Laio, N. Tantalo, M. Parrinello, and R. Petronzio, Biophys J., **9**, 3217-3229 (2006).

34. M.G. Krone, L. Hua, P. Soto, R. Zhou, B.J. Berne, and J.E. Shea, J Am Chem Soc., **130**, 11066-72 (2008).

35. P.H. Nguyen, M.S. Li, G. Stock, J.E. Straub, and D. Thirumalai, Proc Natl Acad Sci U S A., **104**, 111-116 (2007).

36. A. Irbäck, and S. Mitternacht, Proteins, **71**, 207-14 (2008).

37. Y. Lu, P. Derreumaux, Z. Guo, N. Mousseau, and G. Wei, Proteins, **75**, 954-963 (2009).

38. X. Ge, Y. Sun, and F. Ding, Biochim Biophys Acta Biomembr., **1860**, 1687-1697 (2018).

39. Y. Lu, G. Wei, and P. Derreumaux, J Chem Phys., **137**, 025101 (2012).

40. J. Nasica-Labouze, P.H. Nguyen, F. Sterpone, O. Berthoumieu, N.V. Buchete, S. Coté, A. De Simone, A.J. Doig, P. Faller, A. Garcia et al., Chem Rev., **115**, 3518-3563 (2015).

41. J. Nasica-Labouze, M. Meli, P. Derreumaux, G. Colombo, and N. Mousseau, PLoS Comput Biol., **7**, e1002051 (2011).

42. A. Iorio, Š. Timr, L. Chiodo, P. Derreumaux, and F. Sterpone Proteins. doi: 10.1002/prot.26500 (2023).

43. F. Baftizadeh, X. Biarnes, F. Pietrucci, F. Affinito, and A. Laio, J Am Chem Soc., **134**, 3886-3894 (2012).

44. F. Baftizadeh, F. Pietrucci, X. Biarnés, and A. Laio, Phys Rev Lett., **110**, 168103 (2013).

45. L.F. Trugilho, and L. G. Rizzi, EPL ,**137** 57001 (2022).

46. R.I. Dima, and D. Thirumalai, Protein Sci., **11**, 1036-49 (2002).

47 P.M. Harrison, H.S. Chan, S.B. Prusiner, and F.E. Cohen, Protein Sci., **10**, 819-35 (2001).

48. M.S. Li, D.K. Klimov, J.E. Straub, and D. Thirumalai, J Chem Phys., **129**, 175101 (2008).

49. L.F. Trugilho, and L. G. Rizzi, J.: Phys. Conf. Series, **1483,** 012011 (2020).

50. A. Irbäck, and J. Wessén, J. Chem. Phys., **143,** 105104 (2015).

51. M.S. Li, N.T. Co, G. Reddy, C.K. Hu, J.E. Straub, and D. Thirumalai, Phys Rev Lett., **105**, 218101 (2010).

52. F. Sterpone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragnolini, Y.C. Chebaro, J.F. St-Pierre, M. Kalimeri, A. Barducci et al., Chem Soc Rev., **43**, 4871-4893 (2014).

53. S. Abeln, M. Vendruscolo, C.M. Dobson, and D. Frenkel, PLoS One, **9**, e85185 (2014).

54. T.T. Tran, P.H. Nguyen, and P. Derreumaux, J Chem Phys., **144**, 205103 (2016).

55. F. Sterpone, S. Doutreligne, T.T. Tran, S. Melchionna, M. Baaden, P.H. Nguyen, and P. Derreumaux, Biochem Biophys Res Commun., **498**, 296-304 (2018).

56. N Metropolis, A.W. Rosenbluth, M.N Rosenbluth, and A.H. Teller, J Chem Phys., **21**, 1087 (1953).

57. A. Irbäck, S.A.E. Jónsson, N. Linnemann, B. Linse, and S. Wallin, Phys Rev Lett., **110**, 058101 (2013).

58. J. Zhang, and M. Muthukumar, J Chem Phys., **130**, 035102 (2009).

59. P.H. Nguyen, M.S. Li, and P. Derreumaux, Phys Chem Chem Phys., **13**, 9778-88 (2011).

60. V.H. Man, X. He, P. Derreumaux, B. Ji, X.Q. Xie, P.H. Nguyen, and J. Wang, J Chem Theory Comput., **15**, 1440-1452 (2019).

61. W. Hwang, S. Zhang, R.D. Kamm, and M. Karplus, Proc Natl Acad Sci U S A., **101**, 12916-12921 (2004).

62. Y. Zou, Y. Sun, Y. Zhu, B. Ma, R. Nussinov, and Q. Zhang, ACS Chem Neurosci., **7**, 286-296 (2016).

63. P.H. Nguyen, and P. Derreumaux, J Phys Chem B., **117**, 5831-40 (2013).

64. D.W. Li, S. Mohanty, A. Irbäck, and S. Huo, PLoS Comput Biol., **4**, e1000238 (2008).

65. D. Matthes, V. Gapsys, J.T. Brennecke, and B.L. de Groot, Sci Rep., **6**, 33156 (2016).

66. A. Melquiond, G. Boucher, N. Mousseau N, and P. Derreumaux, J Chem Phys., **122**, 174904 (2005).

67. A. Melquiond, N. Mousseau, and P. Derreumaux, Proteins., **65**, 180-191 (2006).

68. G. Wei, N. Mousseau, and P. Derreumaux, Biophys J., **87**, 3648-3656 (2004).

69. S. Patel, Y.U. Sasidhar, and K.V.R Chary, J Phys Chem B. 201, 121, 7536-7549 (2017).

70. I.M. Ilie, and A. Caflisch, Chem Rev., **119**, 6956-6993 (2019).

71. C. Liang, R. Ni, J.E. Smith, W.S. Childers, A.K. Mehta, and D.G. Lynn, J Am Chem Soc., **136**, 15146-9 (2014).

72. X. Li, J. Lei, R. Qi, L. Xie, and G. Wei, Phys Chem Chem Phys., **21**, 15686-15694 (2019).

73. Y. Okada, K. Okubo, K. Ikeda, Y. Yano, M. Hoshino, Y. Hayashi, Y. Kiso, H. Itoh-Watanabe, A. Naito, and K. Matsuzaki, ACS Chem Neurosci., **10**, 563-572 (2019).

74. L. Yu, R. Edalji, J.E. Harlan, T.F. Holzman, A.P. Lopez, B. Labkovsky, H. Hillen, S. Barghorn, U. Ebert, and P.L. Richardson et al., Biochemistry, **48**, 1870-1877 (2009).

75. Z. Jia, J.D. Schmit JD, and J. Chen, Proc Natl Acad Sci U S A., **117**, 10322-10328 (2020).

76. S.A. Petty, and S.M. Decatur, Proc Natl Acad Sci U S A., **102**, 14272-14277 (2005).

77. S.A. Petty, and S.M. Decatur, J Am Chem Soc., **127**, 13488-13489 (2005).

78. S. Santini, G. Wei, N. Mousseau, and P. Derreumaux, Structure, **12**, 1245-1255 (2004).

79 S. Santini, N. Mousseau, and P. Derreumaux, J Am Chem Soc., **126**, 11509-11516 (2004).

80. A.B. Soriaga, S. Sangwan, R. Macdonald, M.R. Sawaya, and D. Eisenberg, J Phys Chem B., **120**, 5810-5816 (2016).

81. C. Liu, M. Zhao, L. Jiang, P.N. Cheng, J. Park, M.R. Sawaya, A. Pensalfini, D. Gou, A.J. Berk, C.G. Glabe, J. Nowick, and D. Eisenberg, Proc Natl Acad Sci U S A., **109**, 20913-20918 (2012).

82. R. Pellarin, P. Schuet, E. Guarnera, and A. Caflisch J. Am. Chem. Soc. 132, 14960-14970 (2010).

83. H. Santuz, P.H. Nguyen, F. Sterpone, and P. Derreumaux, ACS Chem Neurosci., **13**, 711-713 (2022).

84. D. Shea, E. Colasurdo, A. Smith, C. Paschall, S. Jayadev, C.D. Keene, D. Galasko, A. Ko, G. Li, Peskind E. et al., Proc Natl Acad Sci U S A., **119**, e2213157119 (2022).

85. W. Song, G. Wei, N. Mousseau, and P. Derreumaux, J Phys Chem B., **112**, 4410-4418 (2008).

86. P.H. Nguyen, F. Sterpone, and P. Derreumaux, J Phys Chem B., **126**, 10317-10326 (2022).

87. M. Zhu, L. Zeng, Z. Li, C. Wang, L. Wu, and X. Jiang, ChemistryOpen., **12**, e202200253 (2023).

88. A.J. Doig, M.P. Del Castillo-Frias, O. Berthoumieu, B. Tarus, J. Nasica-Labouze, F. Sterpone, P.H. Nguyen, N.M. Hooper, P. Faller, and P. Derreumaux, ACS Chem Neurosci., **8**, 1435-1437 (2017).

89. D.M. Walsh, and D.J. Selkoe, Curr Opin Neurobiol., **61**, 116-124 (2020).

90. B. Jiang, U. Umezaki, A. Augustine, V.M. Jayasinghe-Arachchige, L.F. Serafim, Z.M.S. He, K.M. Wyss, R. Prabhakar, and A.A. Martí, Chem Sci., **14**, 1072-1081 (2023).

91. T.C.T Michaels, A. Šarić, G. Meis, G.T. Heller, S. Curk, P. Arosio, S. Linse, C.M. Dobson, M. Vendruscolo, and T.P.J. Knowles, Proc Natl Acad Sci U S A., **117**, 24251-24257 (2020).

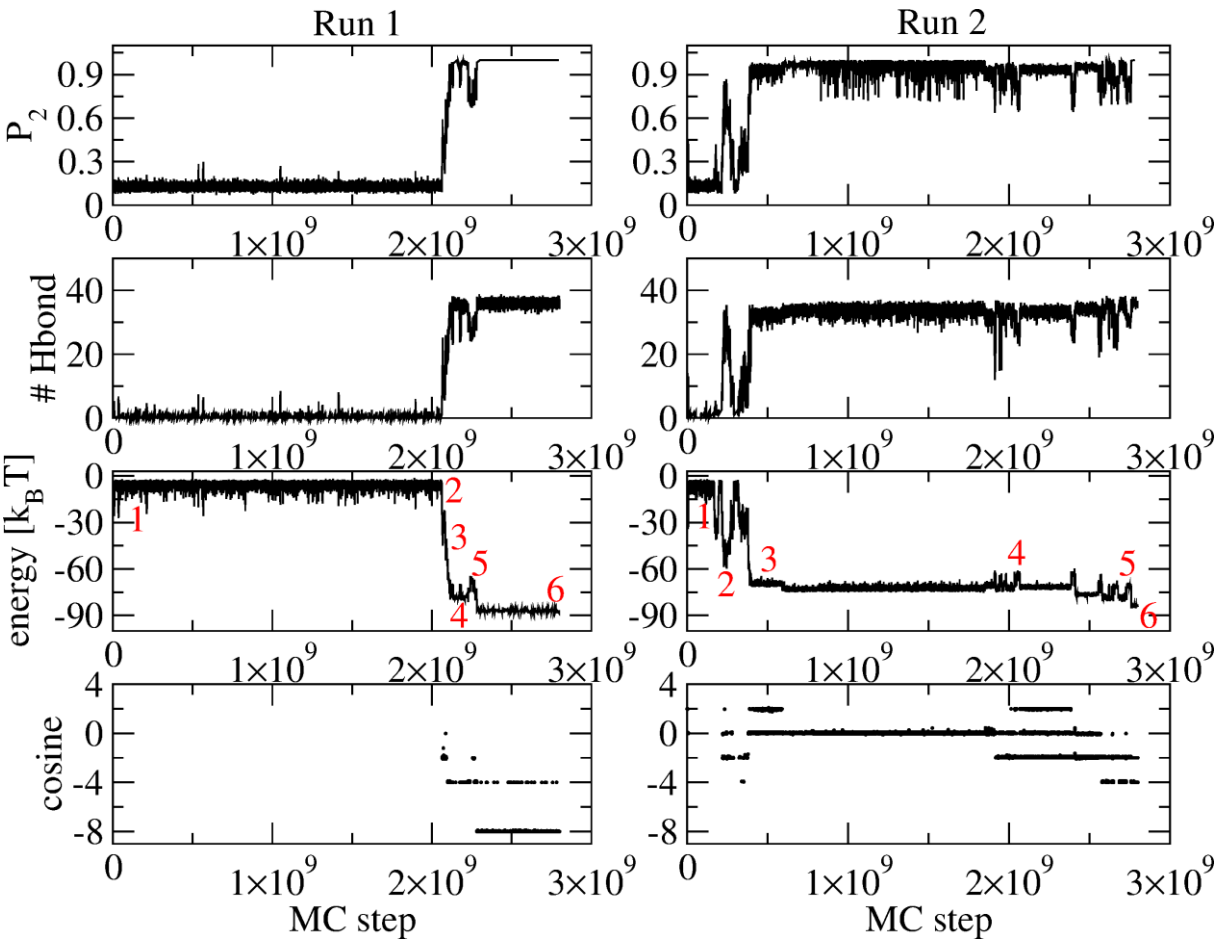


FIG 1. Runs R1 and R2. P2, total number of H-bonds (# Hbond), total energy in k_BT, and the cosine angle of five stranded β-strands in beta-sheets as a function of MC steps up to 2.8 10^9 obtained at T* = 0.27. The states described by numbers in the evolution of energy are described in Figures 3 and 4.

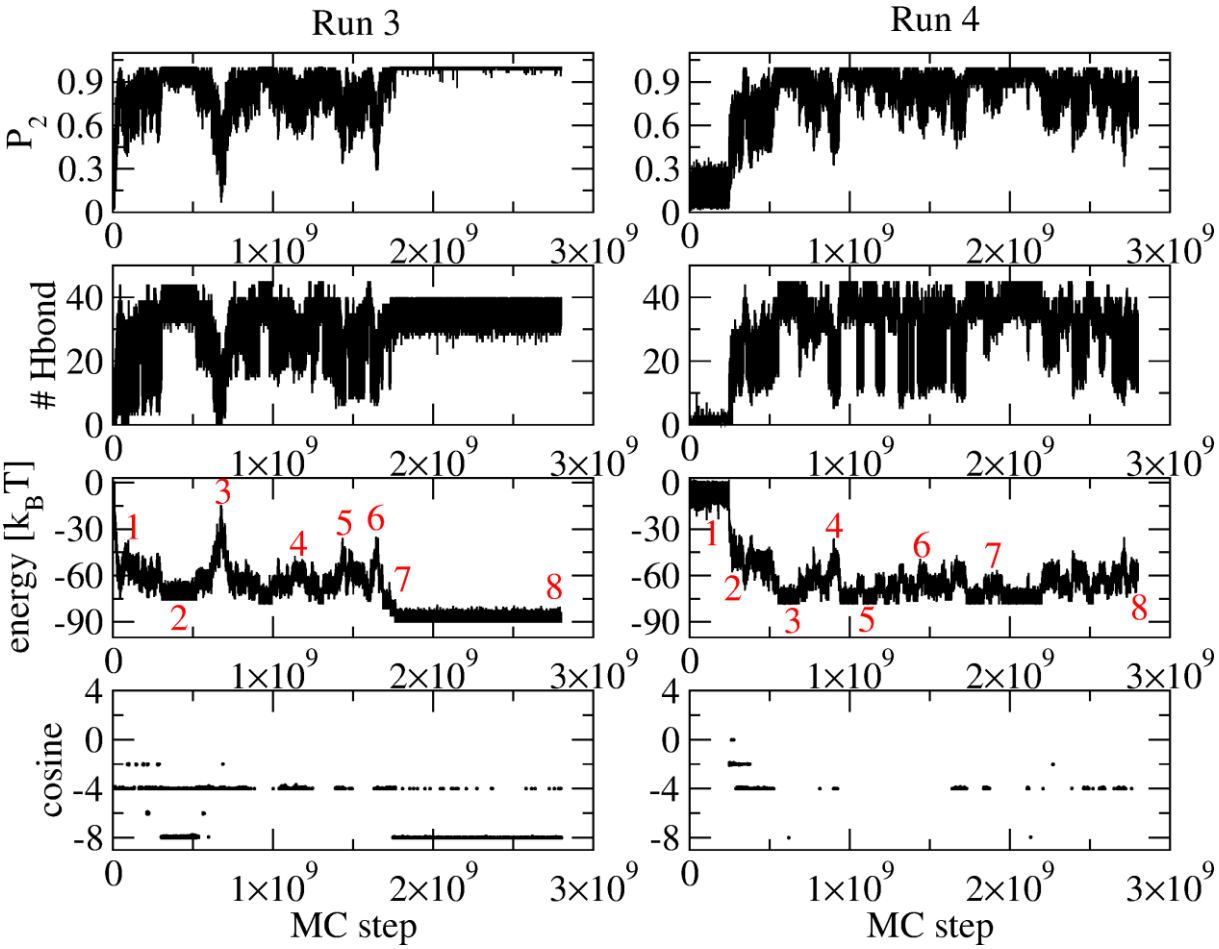


FIG 2. Runs R3 and R4. P2, # Hbond, total energy, and the cosine angle of five stranded beta-strands in beta-sheets as a function of MC steps up to 2.8 10^9 obtained at T* = 0.27. The states described by numbers in the evolution of energy are described in Figures 4 and 5.

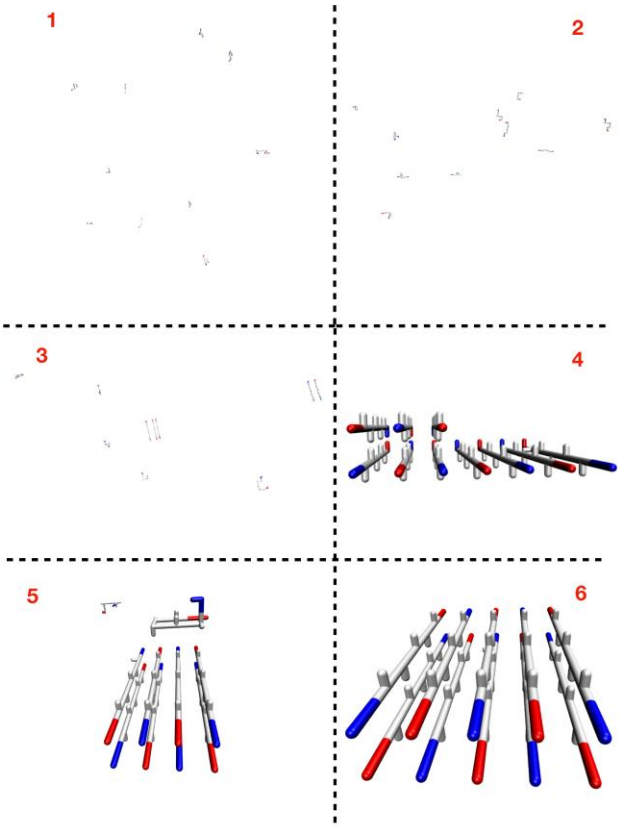


FIG 3. Conformations at six given MC steps in Run 1 shown in Figure 1. The residue K16 is in blue, residue D22 in red, and residues L17, V18, F19, F20 and A21 are in grey. The same colour coding of the amino acids is used in the next figures.

Accepted to J. Chem. Phys. 10.1063/5.0149419

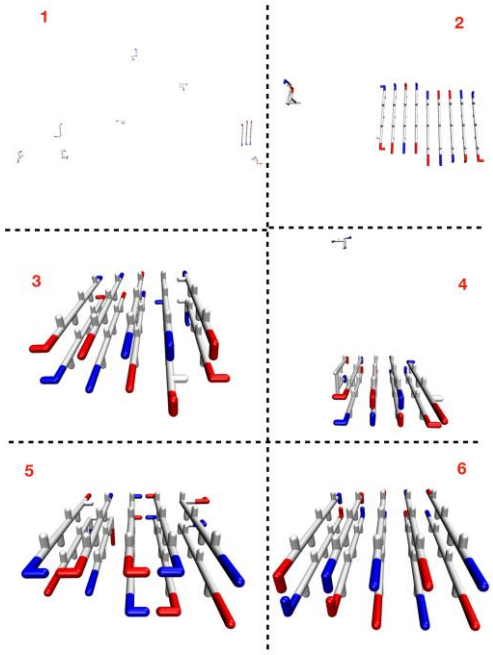


FIG 4. Conformations at six given MC steps of Run 2 shown in Figure 1.

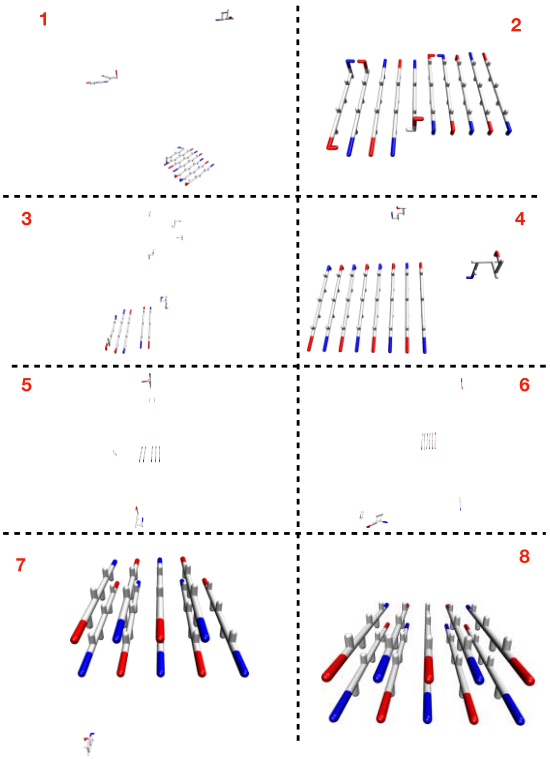


FIG 5. Conformations at given 8 MC steps in Run 3 shown in Figure 2.

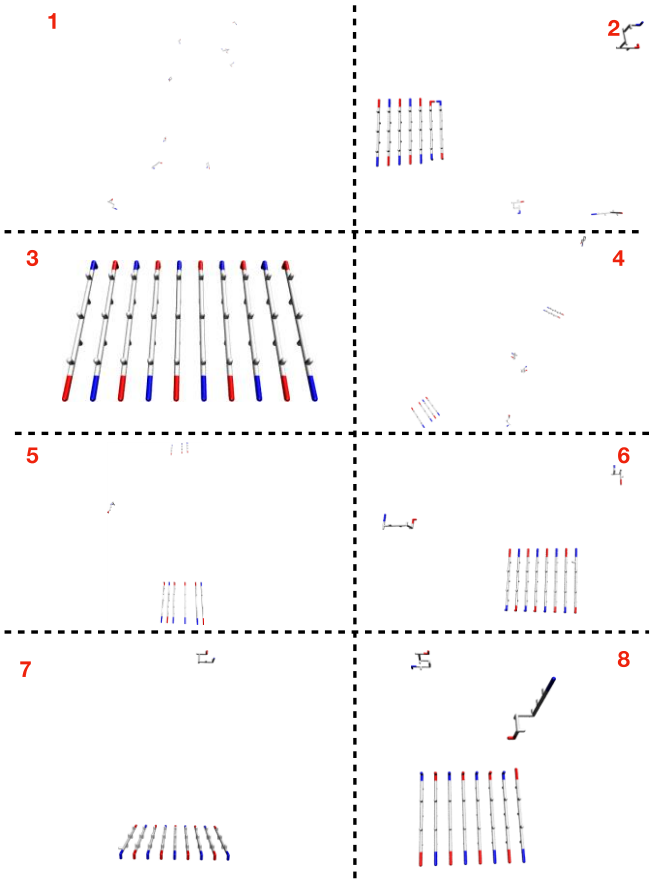


FIG 6. Conformations at given 8 MC steps in Run 4 shown in Figure 2.

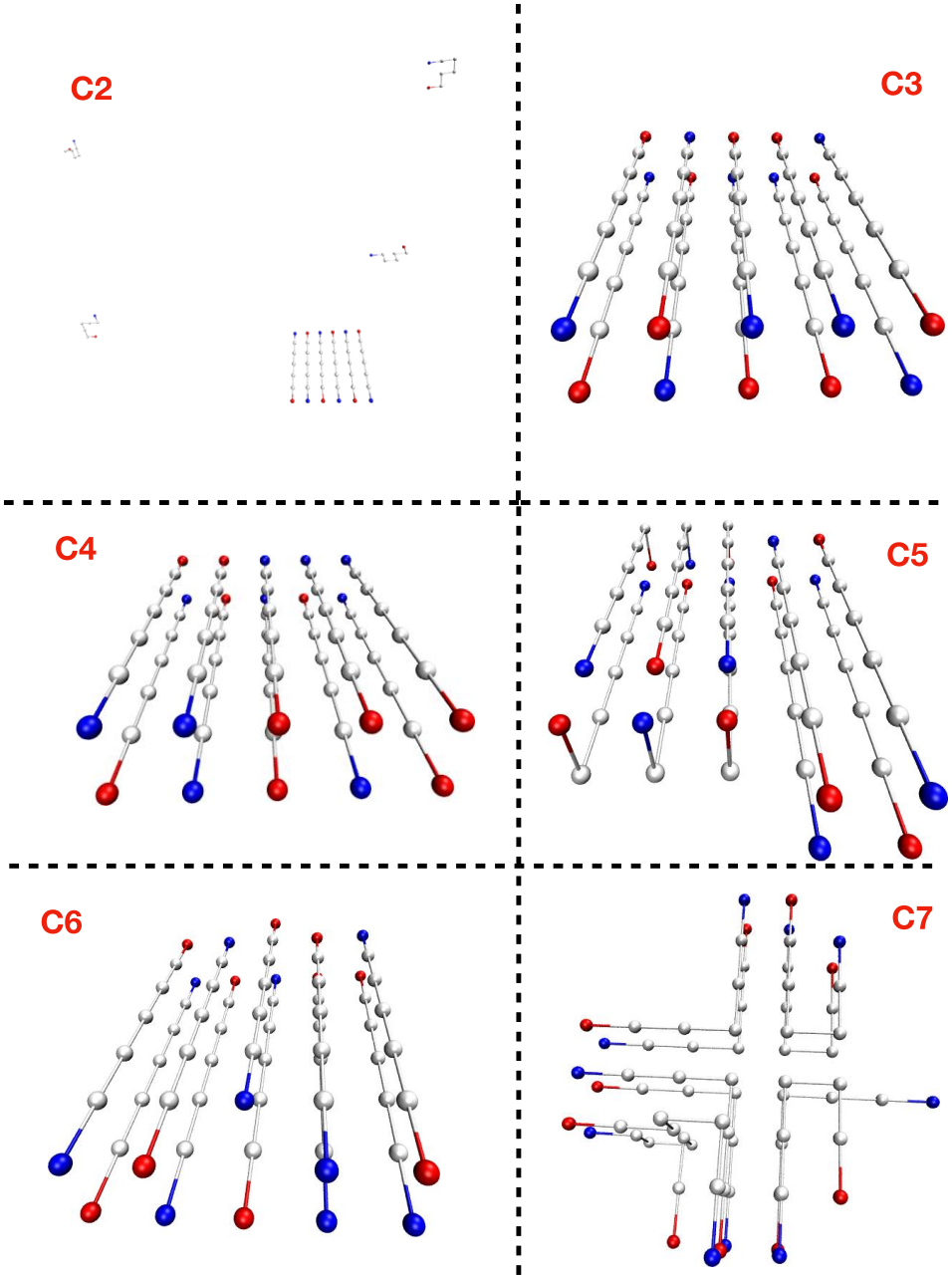


FIG 7. Representative final structures of the six dominant clusters C2-C7 observed in the 41 non-converging simulations of 2.8 10^9 MC steps.

Accepted to J. Chem. Phys. 10.1063/5.0149419

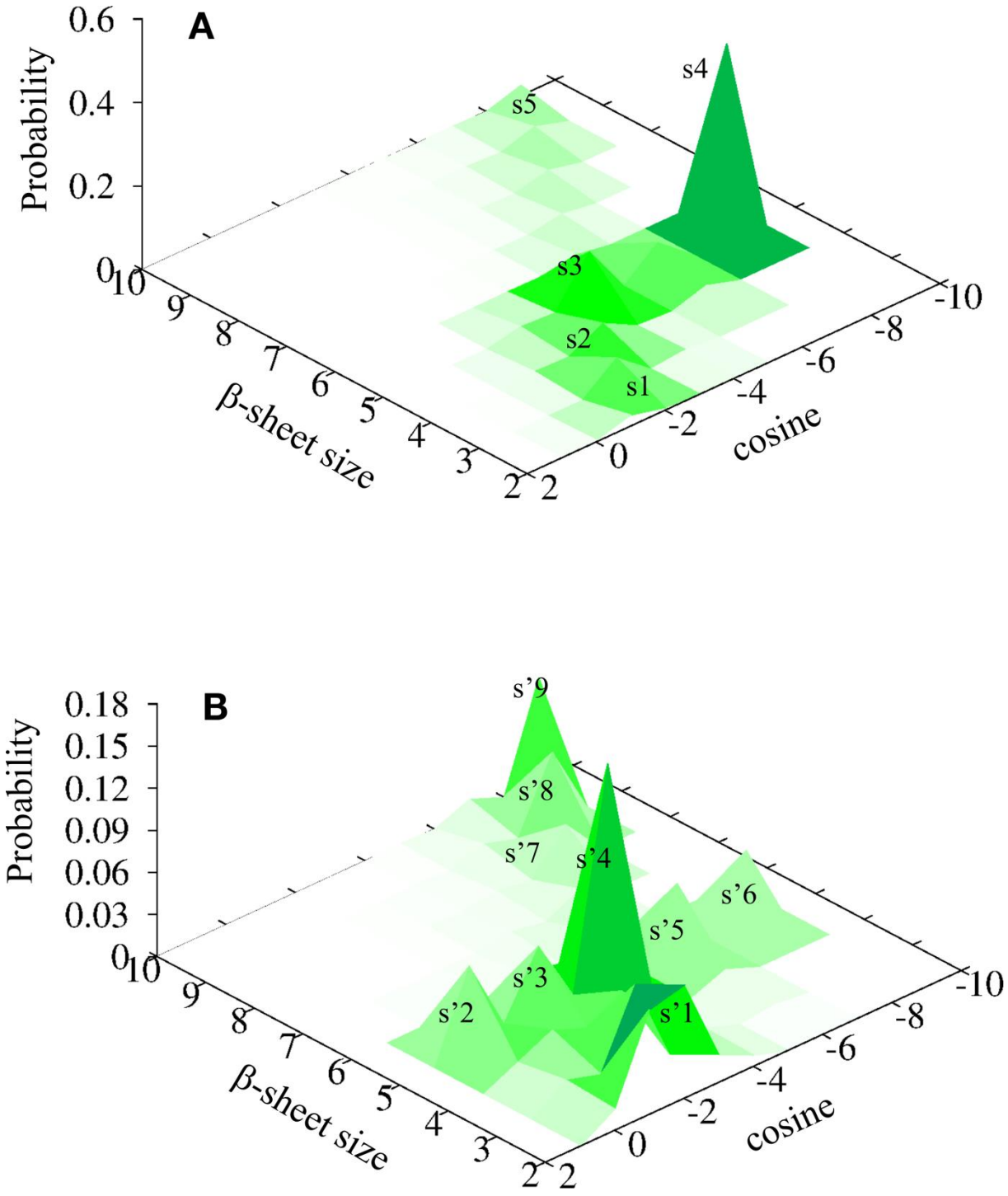


FIG 8. Probability distributions P(cosine angle, β-sheet size) obtained from the 24 converging (A) and 41 non-converging (B) simulations. Shown are the probabilities varying from 0 to 1 of the states using full trajectories.

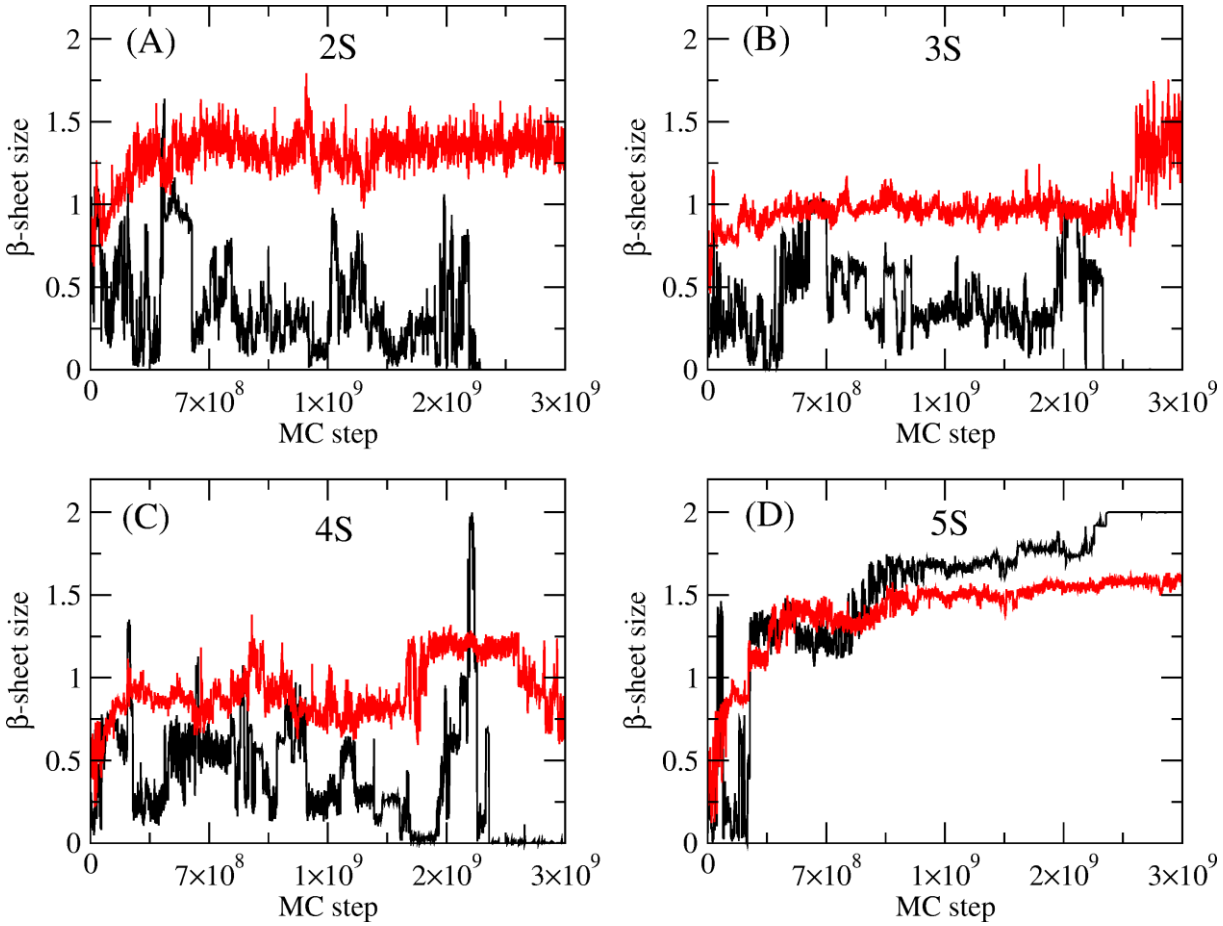


FIG 9. MC evolution of β-sheet sizes. Converging and non-converging simulations are in black and red, respectively. NS stands for one layer made of N-stranded β-strands.

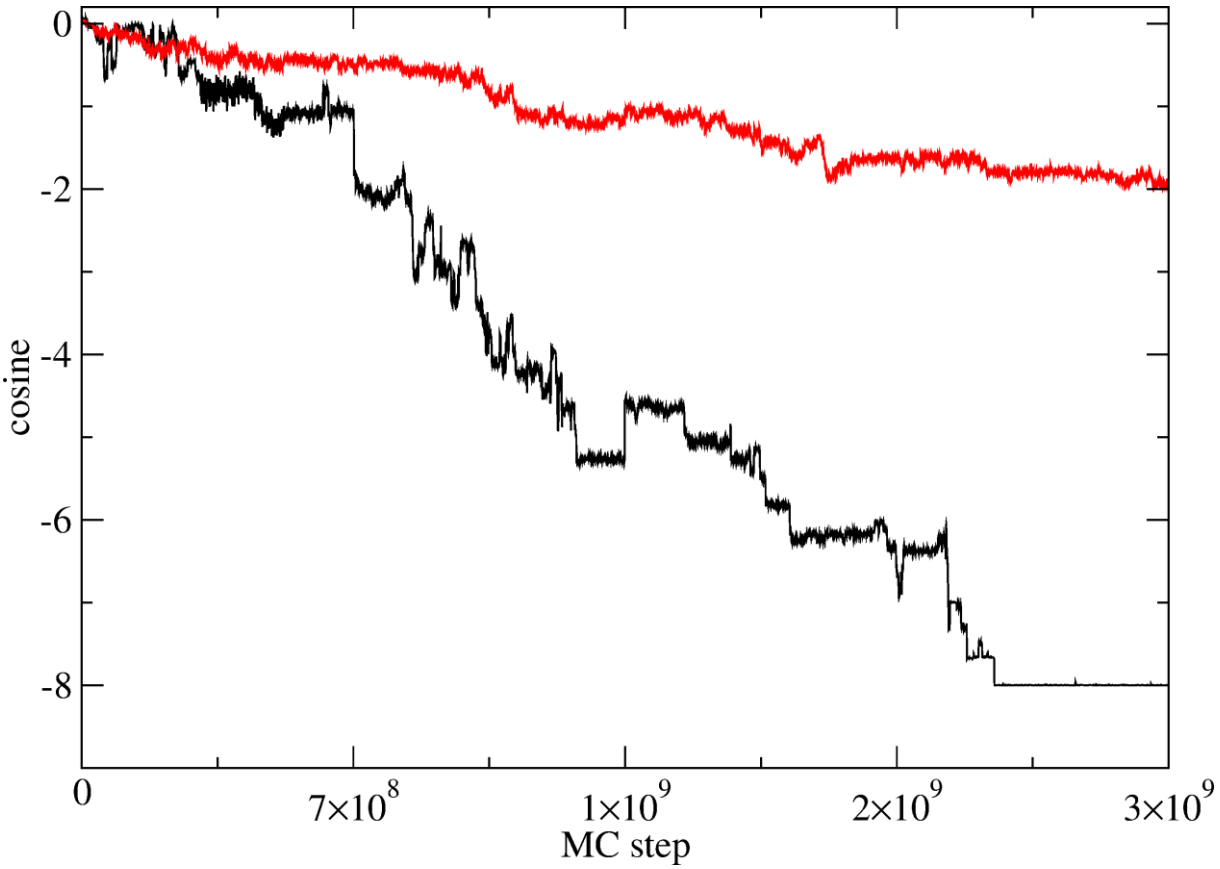
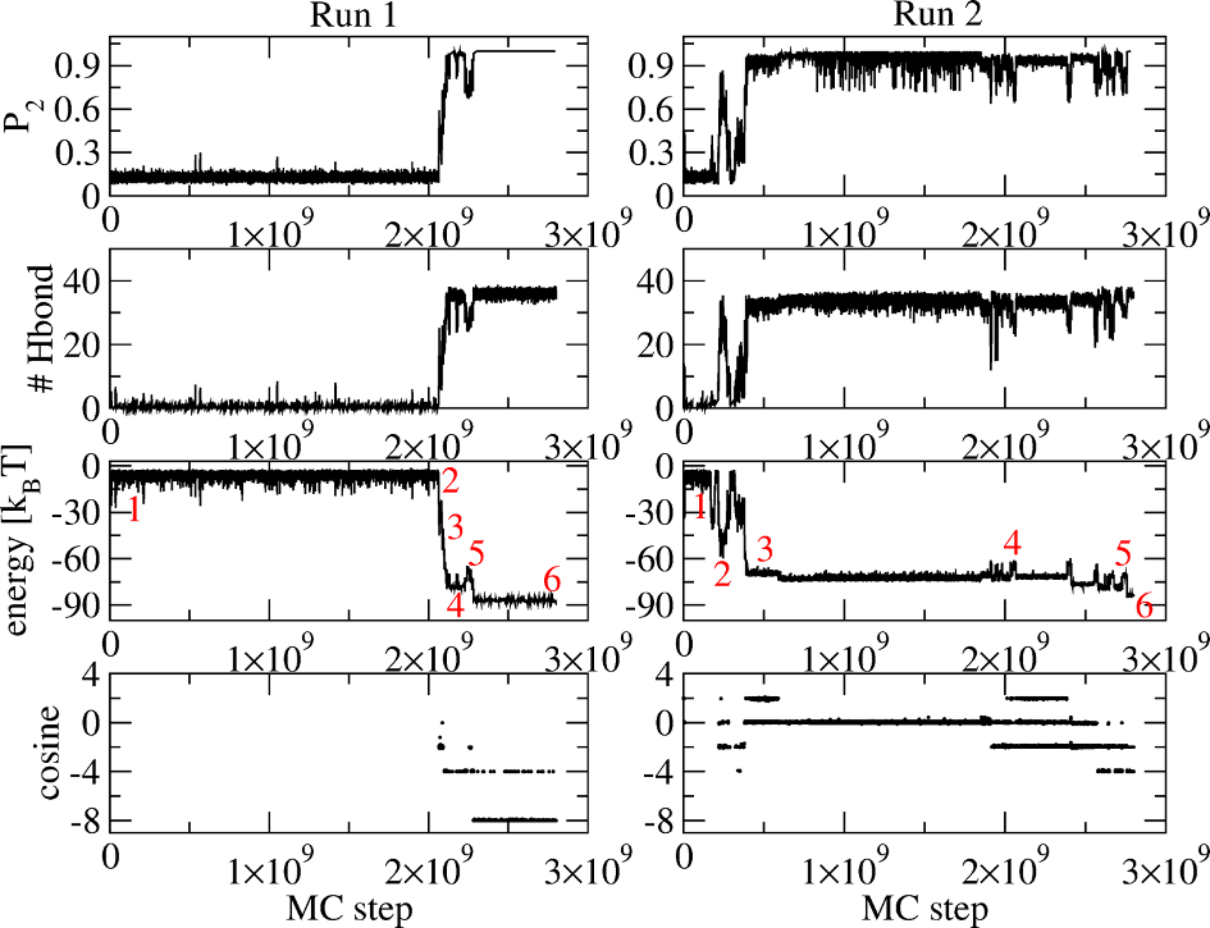
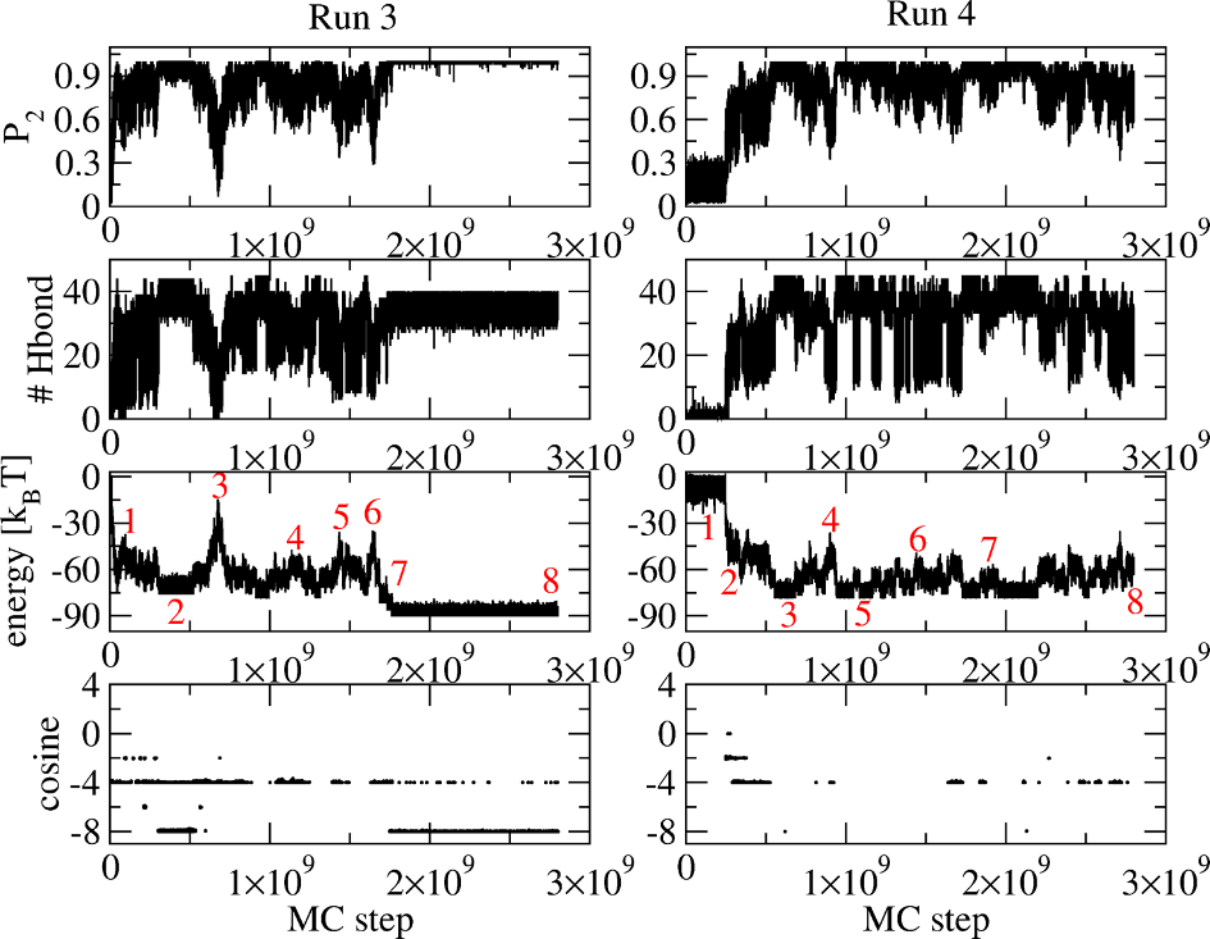


FIG 10. MC evolution of the cosine angle averaged over all converging (black) and non-converging simulations (red).

Trajectory	State	Probability	Cosine	β-sheet size	Anti-parallel β-strand	β-sheet number
	s1	0.065±0.016	-2	3	Y	1
	s2	0.048±0.014	-3	4	Y	1
Converging	s3	0.120±0.064	-4	5	Y	1
	s4	0.467±0.157	-8	5	Y	2
	s5	0.024±0.014	-8	9	Y	1
	s'1	0.080±0.012	-2	2	Y	2
	s'2	0.061±0.011	0	5	N	1
	s'3	0.053±0.009	-2	5	N	1
	s'4	0.161±0.009	-4	5	Y	1
Non-converging	s'5	0.052±0.015	-6	5	N	2
	s'6	0.053±0.013	-8	5	Y	2
	s'7	0.028±0.007	-7	8	Y	1
	s'8	0.051±0.002	-8	9	Y	1
	s'9	0.073±0.035	-9	10	Y	1

Table 1. Topological description of the dominant states identified in all converging and non-converging trajectories shown in Figure 8. For each state, we give the probability (mean and standard deviation), the total cosine angle value, the β-sheet size, whether the strands in the sheets are fully antiparallel (Yes or No), and the number of β-sheet layers. Standard deviation was calculated using two blocks of 12 converging and 20 non-converging simulations.


Run 1
Run 2
P2
Hbond
energy [kBT]
cosine
MC step
MC step
0
0.3
0.6
0.9
20
40
-30
-60
-90
4
-4
-8
1×10^9
2×10^9
3×10^9
1
2
3
4
5
6

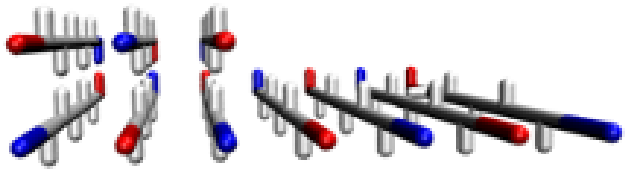

Run 3
Run 4
P2
Hbond
energy [kBT]
cosine
MC step
MC step
0
0.3
0.6
0.9
20
40
-30
-60
-90
4
-4
-8
1×10^9
2×10^9
3×10^9
1
2
3
4
5
6
7
8

1

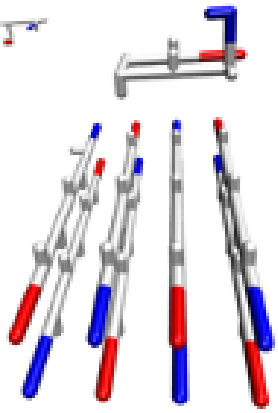
2

3

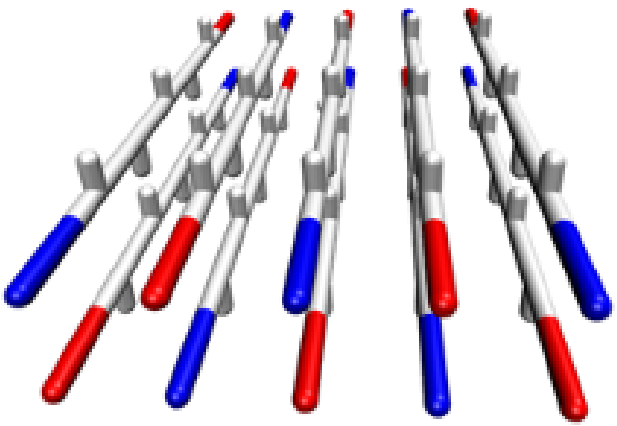
4

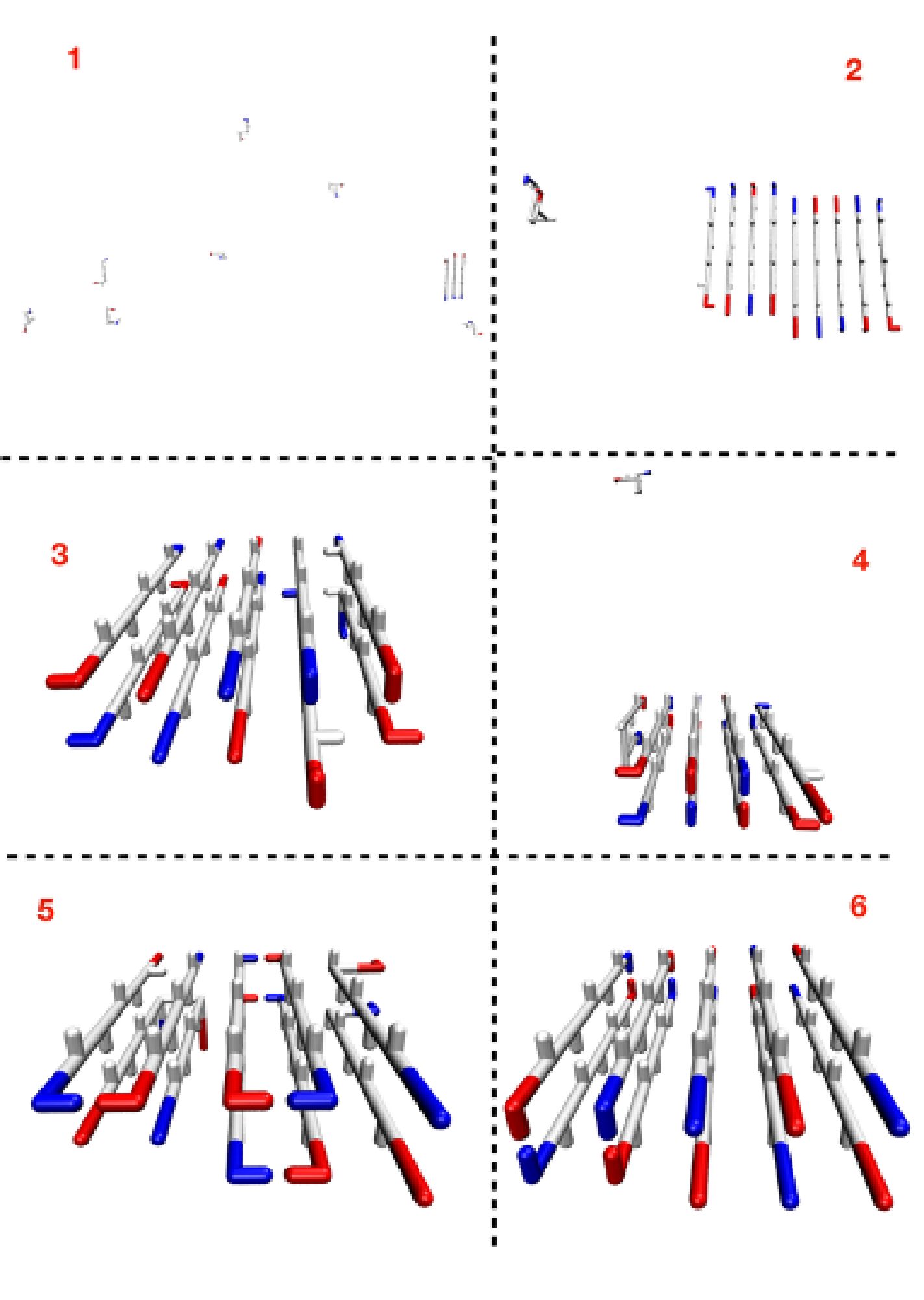


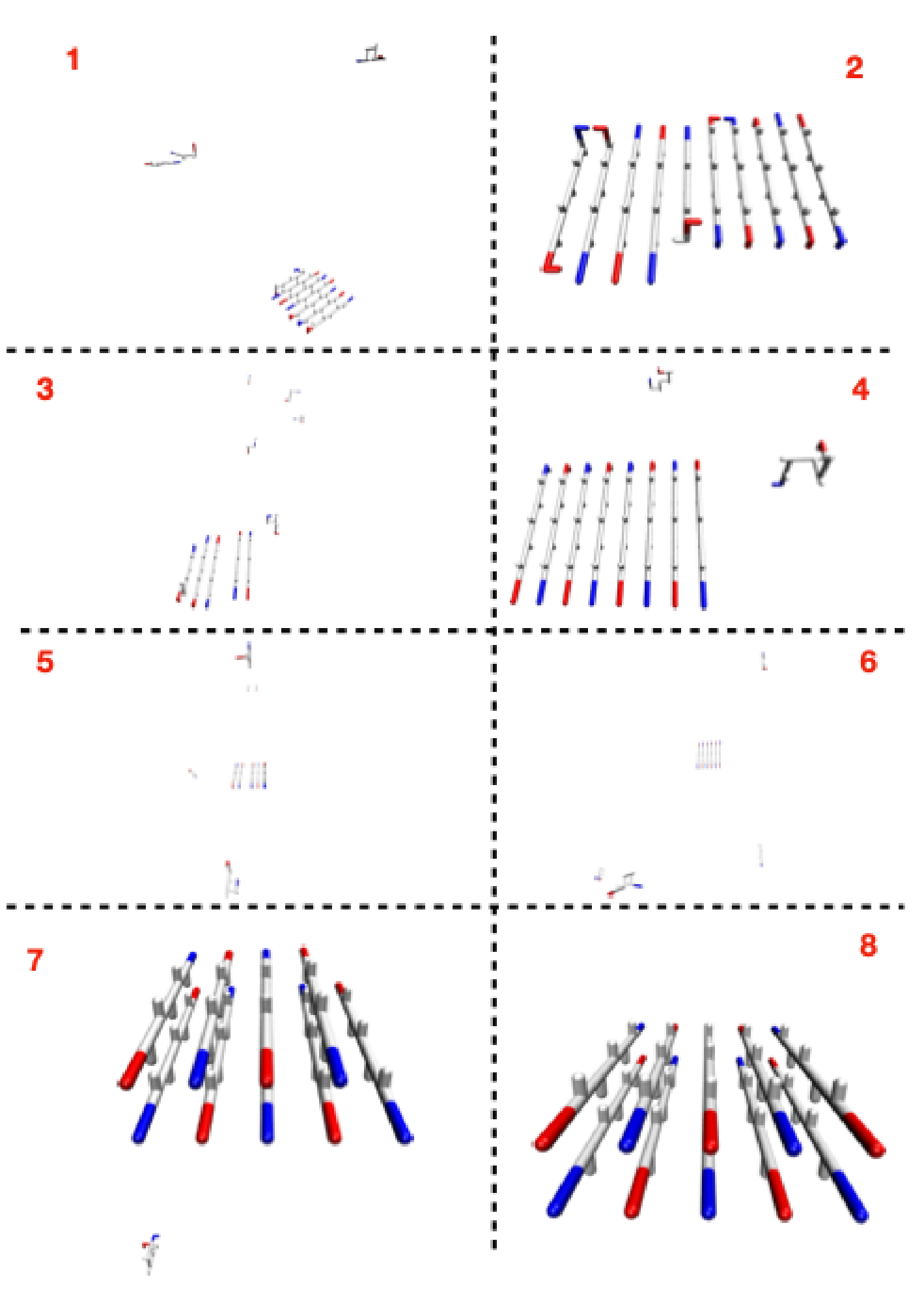
5

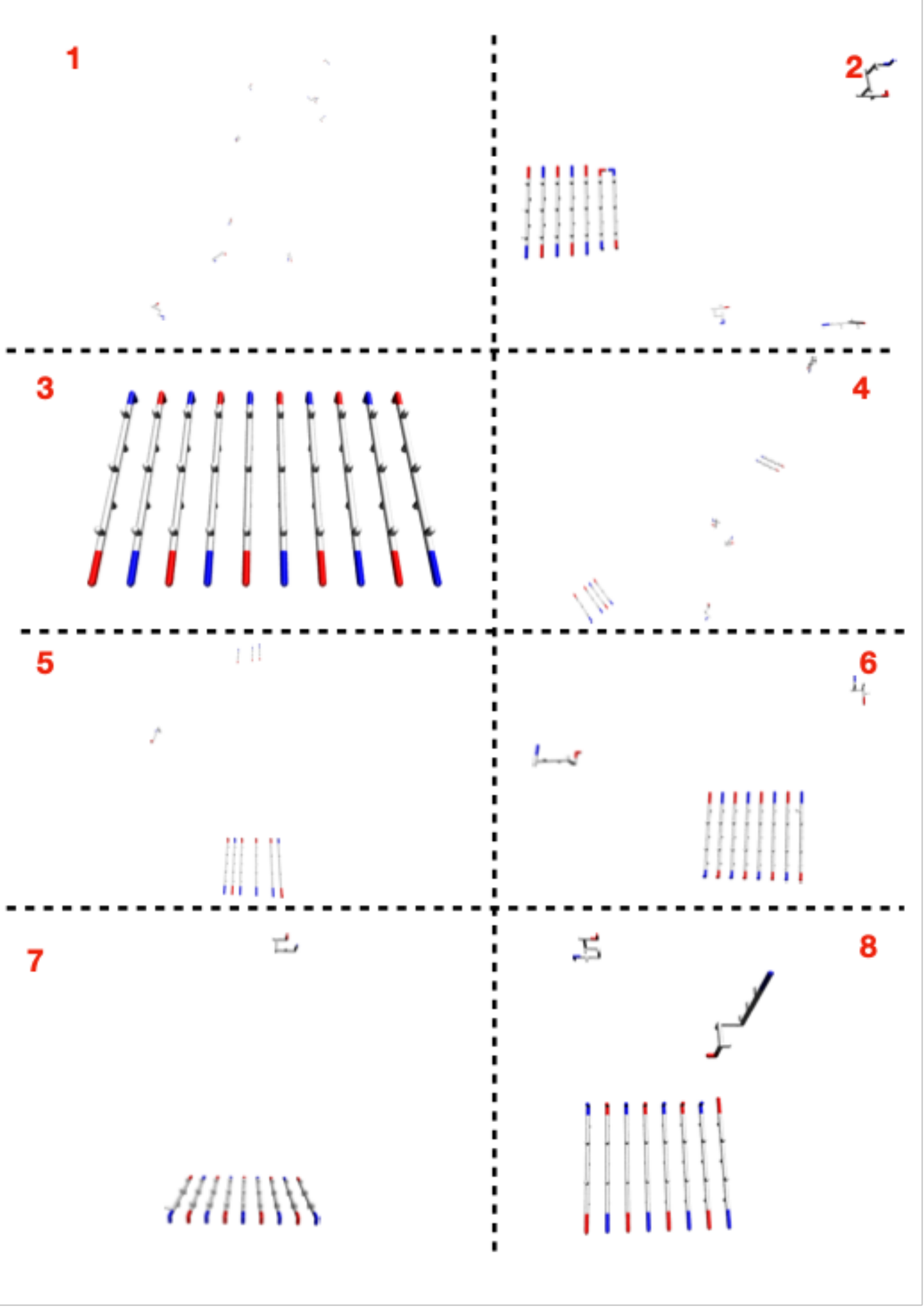


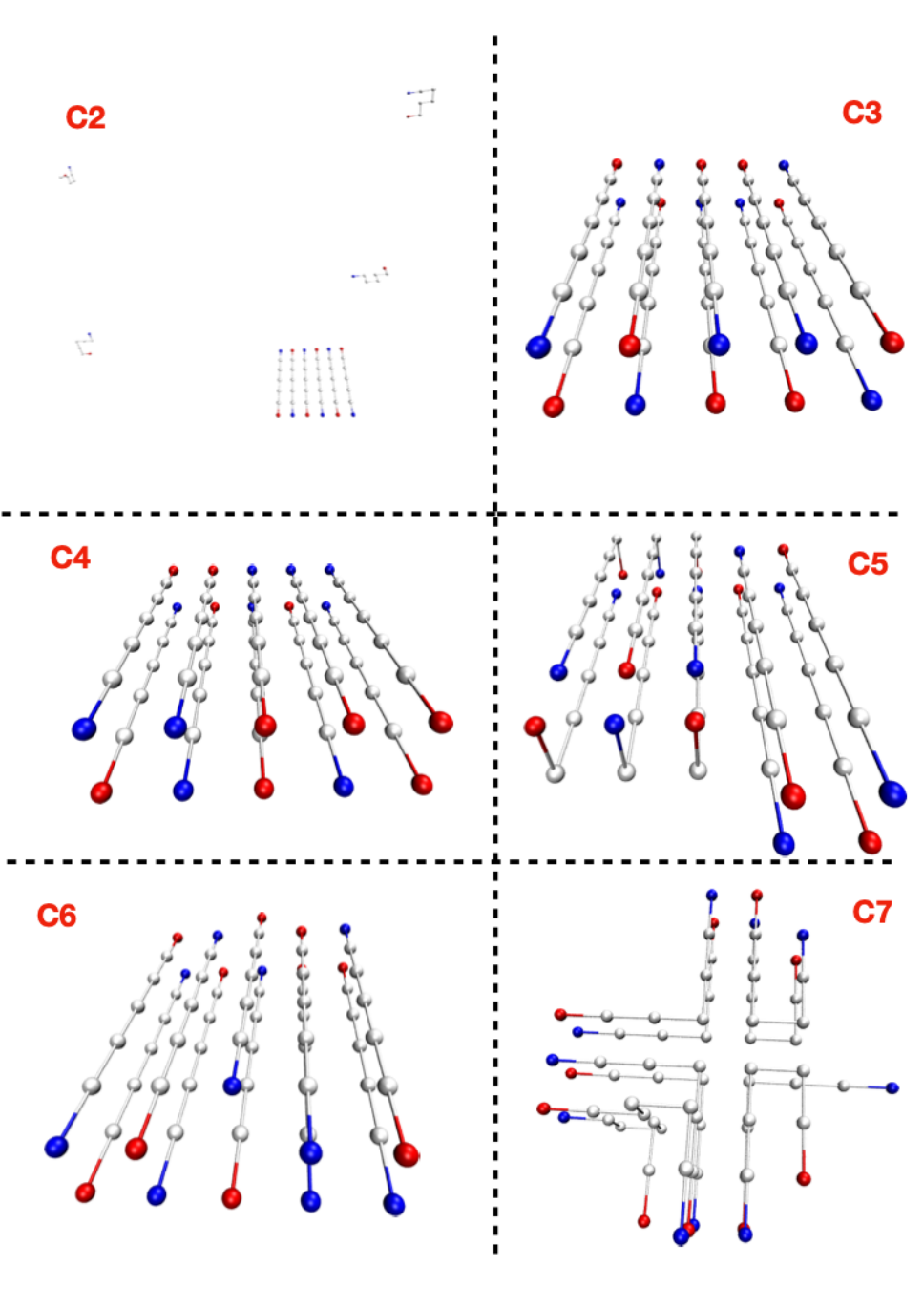
6

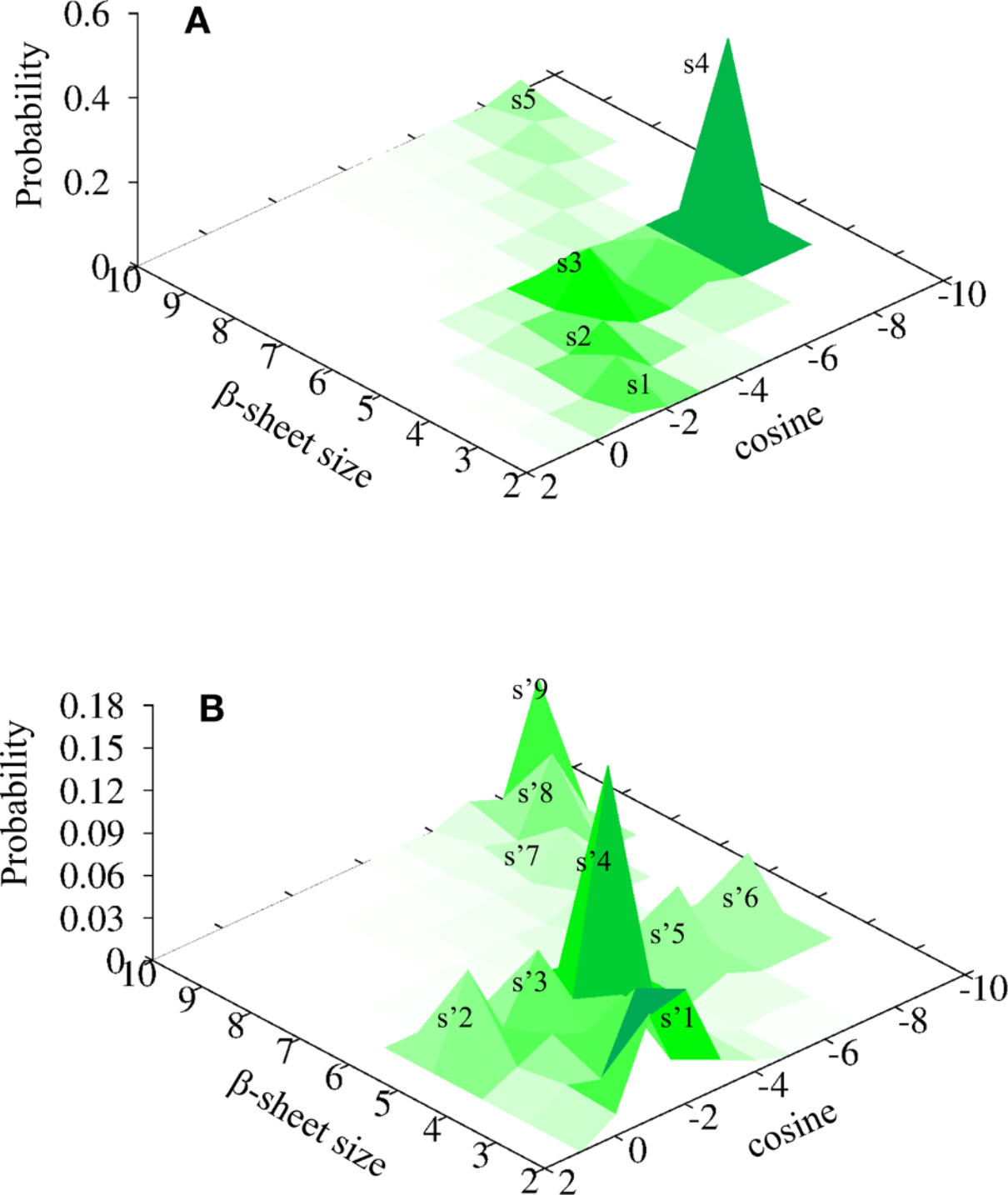


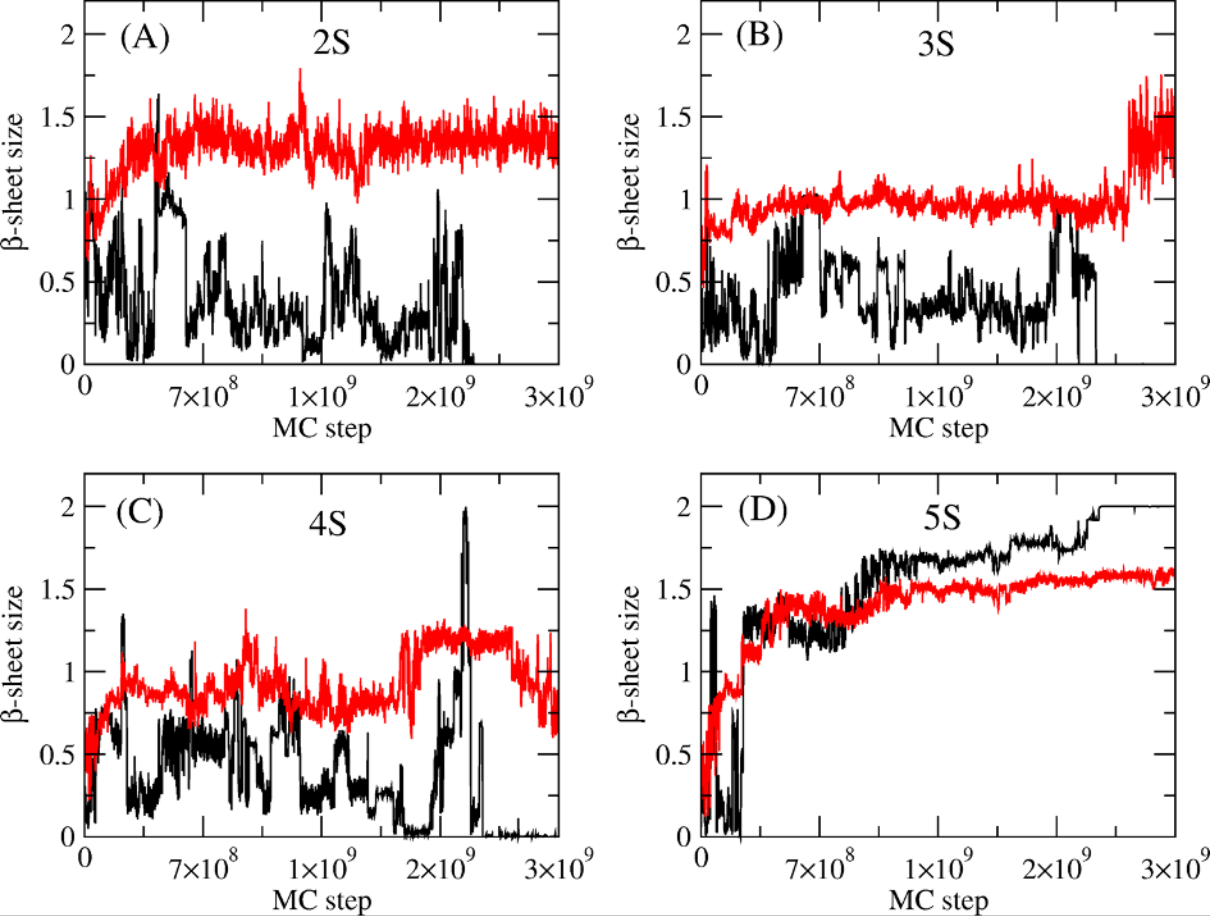

1
2
3
4
5
6

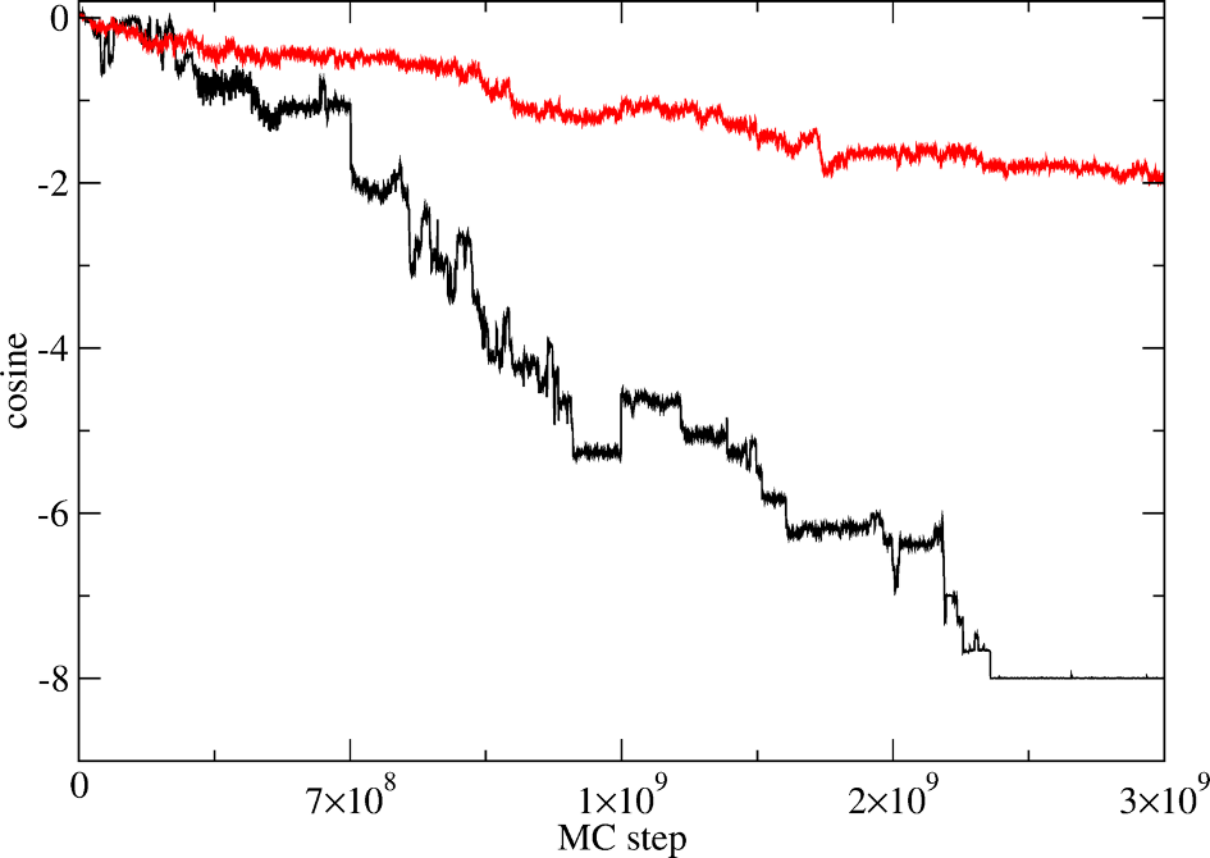

1
2
3
4
5
6
7
8


1
2
3
4
5
6
7
8


C2
C3
C4
C5
C6
C7


A
Probability
0.6
0.4
0.2
0
s5
s4
s3
s2
s1
10
9
8
7
6
5
4
3
2
β-sheet size
2
0
-2
-4
-6
-8
-10
cosine
B
Probability
0.18
0.15
0.12
0.09
0.06
0.03
0
s'9
s'8
s'7
s'4
s'6
s'5
s'3
s'2
s'1
10
9
8
7
6
5
4
3
2
β-sheet size
2
0
-2
-4
-6
-8
-10
cosine


(A)
2S
(B)
3S
(C)
4S
(D)
5S
β-sheet size
MC step
0
0.5
1
1.5
2
7×10^8
1×10^9
2×10^9
3×10^9


0
-2
-4
-6
-8
cosine
0
7×10^8
1×10^9
2×10^9
3×10^9
MC step