Electronic Supplementary Information

Thermodynamics and kinetics of the amyloid-β peptide revealed by Markov state models based on MD data in agreement with experiment

Arghadwip Paul[1,2,‡], Suman Samantray[1,3,‡], Marco Anteghini[1,4], Mohammed Khaled[1], and Birgit Strodel[1,5,*]

[1]*Institute of Biological Information Processing: Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52428 Jülich, Germany,*
[2]*German Research School for Simulation Sciences, RWTH Aachen University, 52062 Aachen, Germany,*
[2]*AICES Graduate School, RWTH Aachen University, Schinkelstraße 2, 52062 Aachen, Germany*
[4]*Current address: LifeGlimmer GmbH, Markelstraße 38 12163 Berlin, Germany,*
[5]*Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany,*
[‡]*These authors contributed equally.*
[*]*Email: b.strodel@fz-juelich.de*

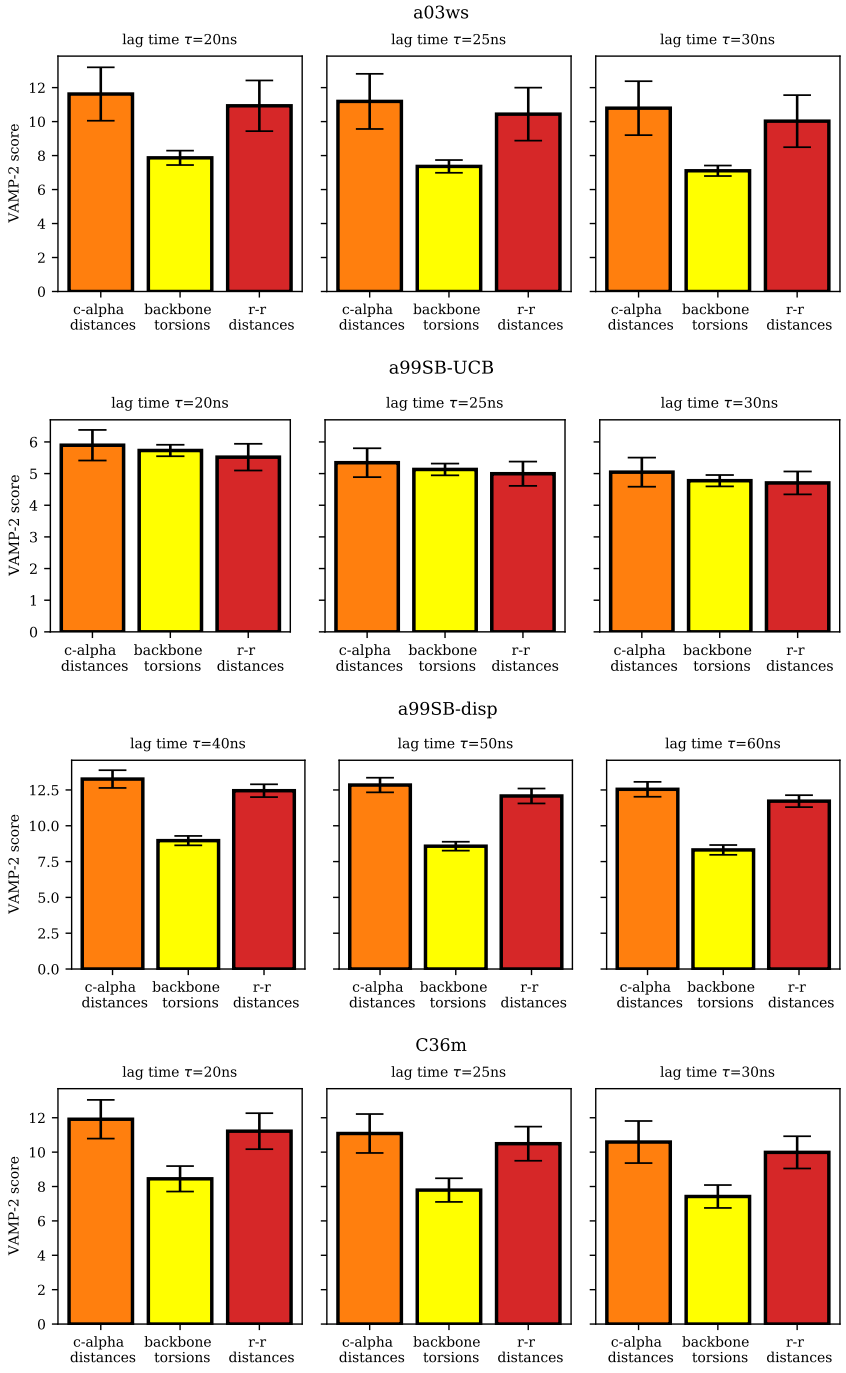

Figure S1: VAMP-2 scores for four force fields (a03ws, a99SB-UCB, a99SB-disp, C36m), three different features (C_α distances, backbone torsion angles, residue–residue distances), and three different lag times. The choice of these lag times was based on the lag times used for building the final MSMs: $\tau = 25$ ns for a03ws, a99SB-UCB, and C36m and $\tau = 50$ ns for a99SB-disp. The scores are the highest for the C_α distances in all the cases (the higher, the better), while they are the smallest for the backbone torsion angles for almost all combinations. We thus chose the C_α distances as the feature for the MSM analysis.

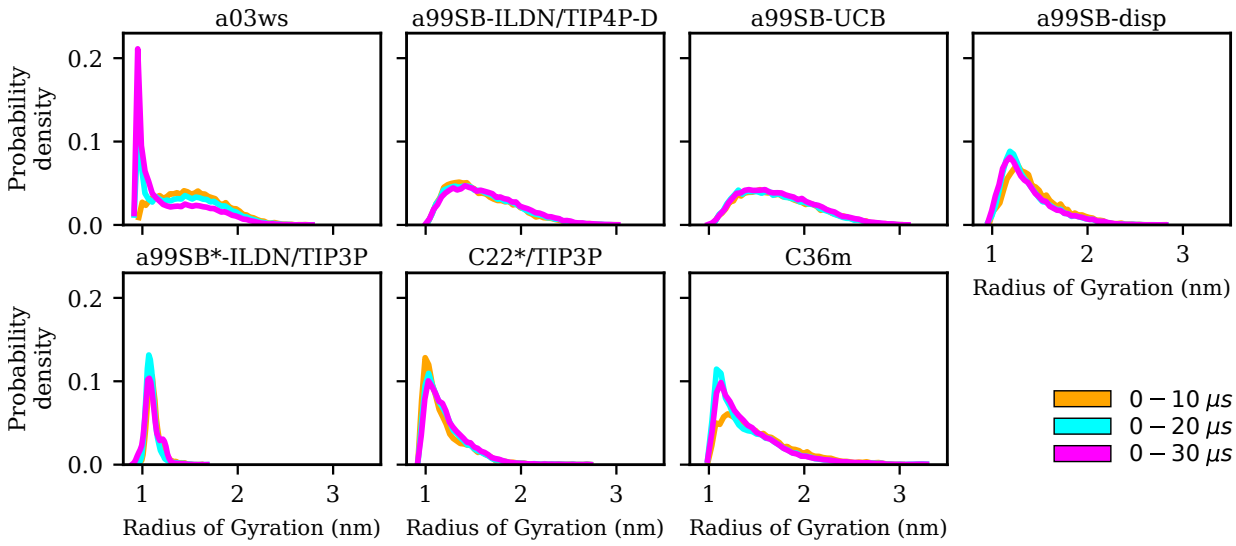

Figure S2: Distribution of the radius of gyration R_{gyr} for increasing trajectory lengths (0–10 µs: yellow, 0–20 µs: cyan, 0–30 µs: magenta) for the different force fields (labels on the top of the panels).

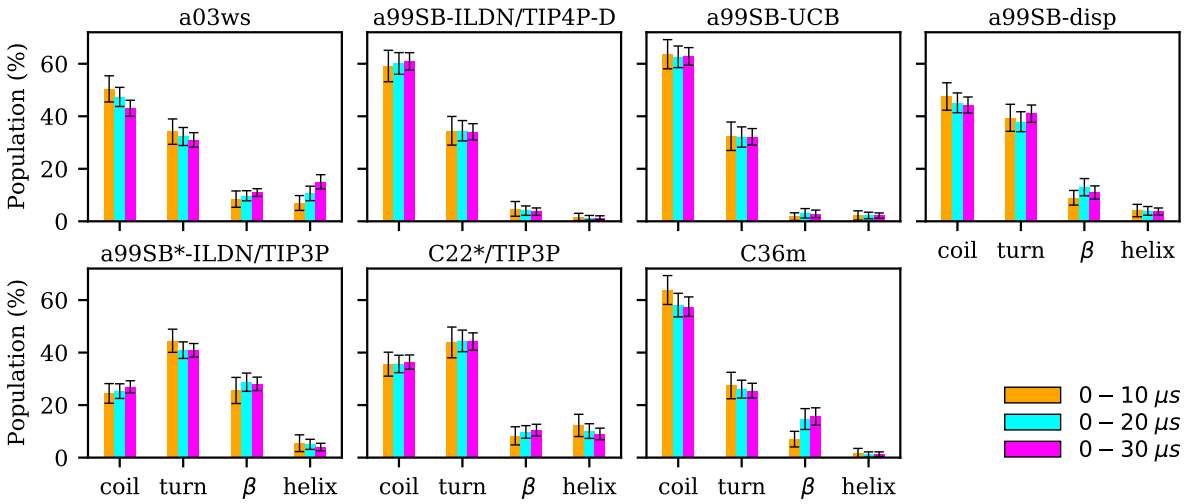

Figure S3: Time-averaged secondary structures coil, turn, β-sheet, and α-helix for increasing trajectory lengths (0–10 µs: yellow, 0–20 µs: cyan, 0–30 µs: magenta) along with standard errors (shown as black bars) for the different force fields (labels on the top of the panels).

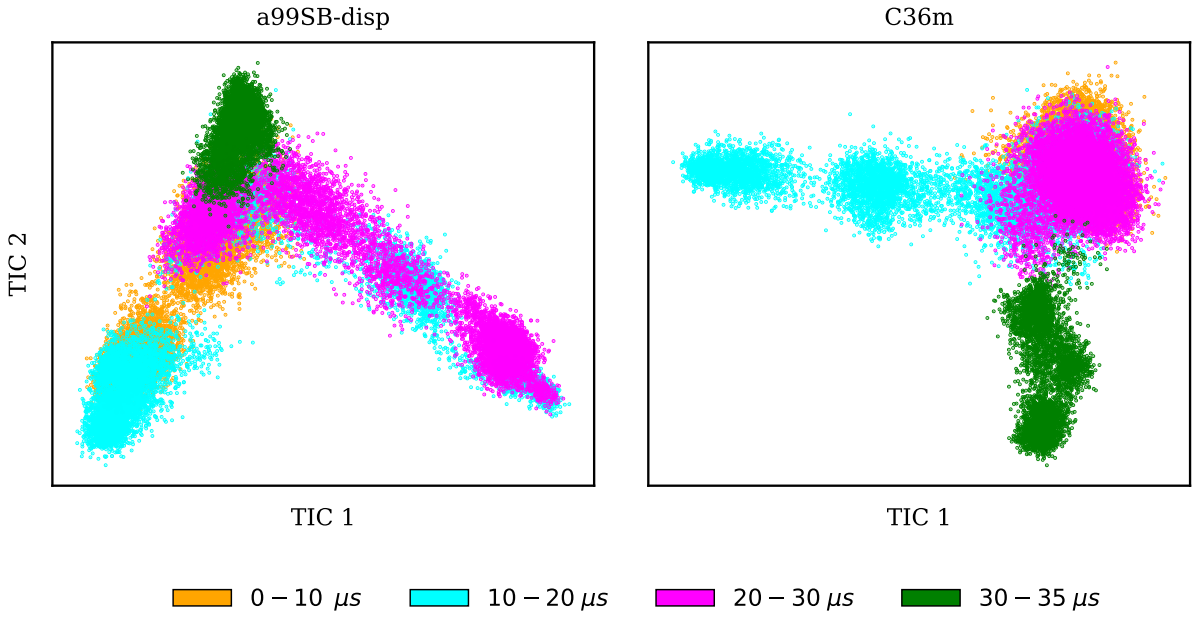

Figure S4: Sample densities for different time windows of the trajectories (0–10 µs: yellow, 10–20 µs: cyan, 20–30 µs: magenta, 30–35 µs: green) projected along the first two TICA components for a99SB-disp (left) and C36m (right).

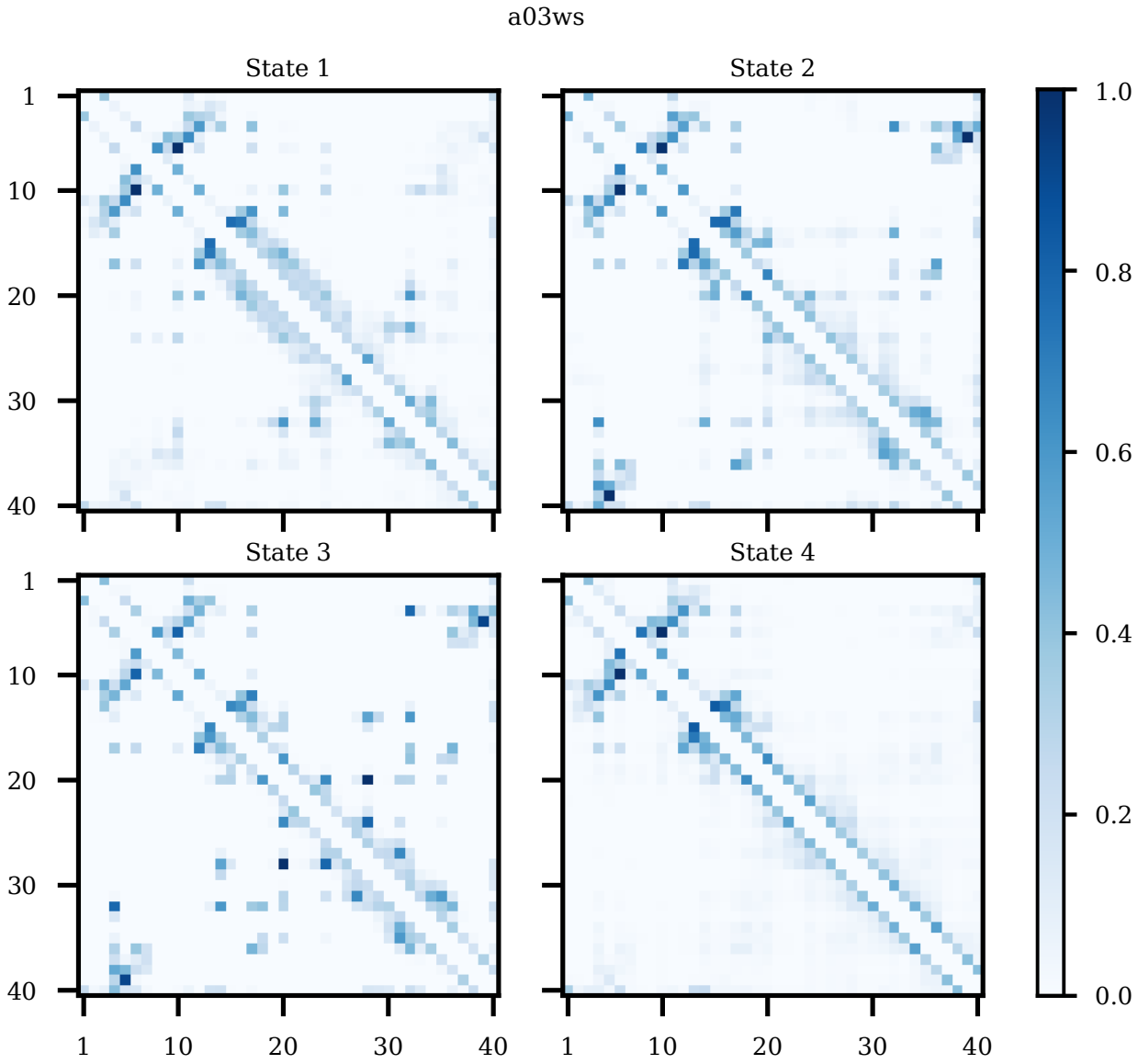

Figure S5: Normalized contacts for the coarse-grained MSM states obtained from the simulation with a03ws.

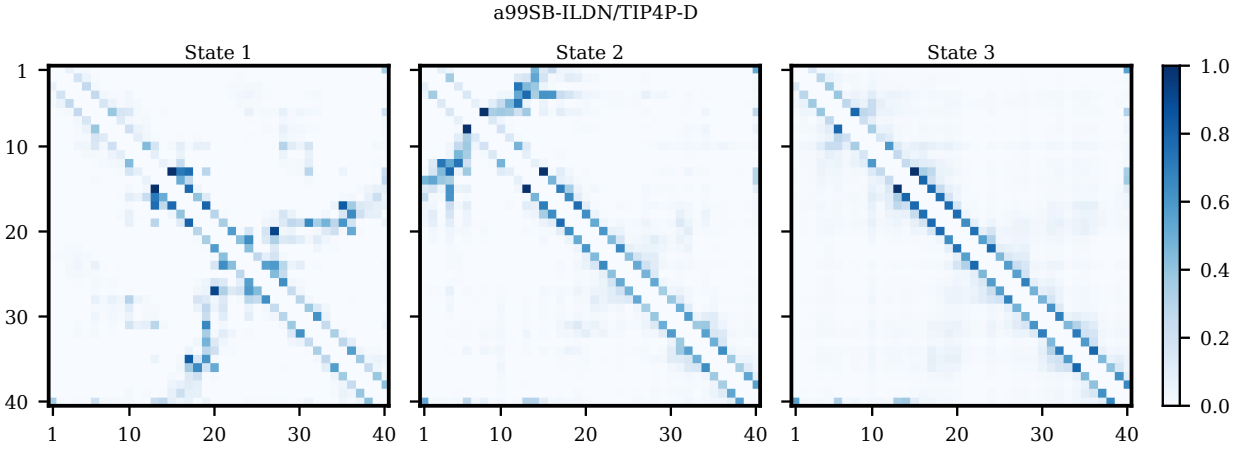

Figure S6: Normalized contacts for the coarse-grained MSM states obtained from the simulation with a99SB-ILDN/TIP4P-D.

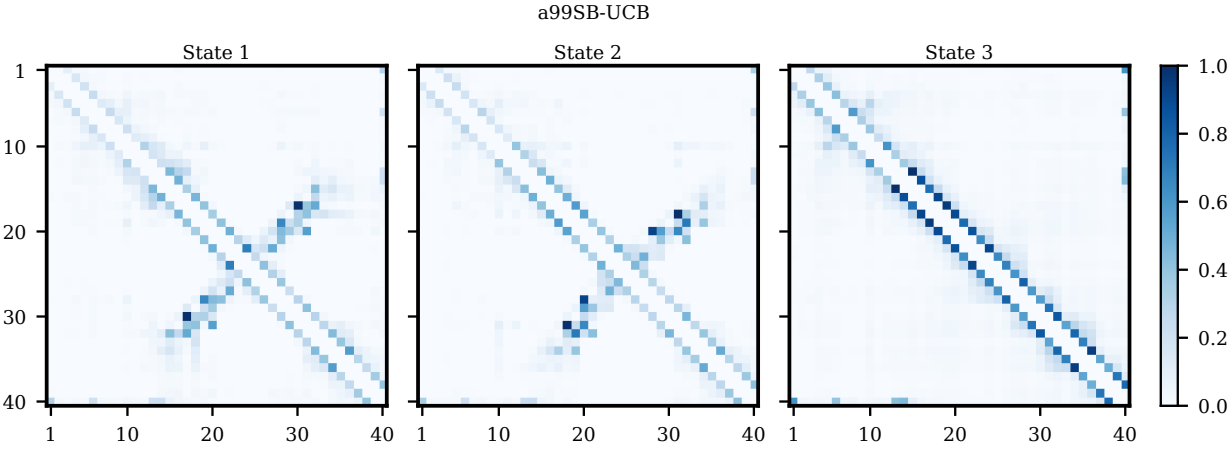

Figure S7: Normalized contacts for the coarse-grained MSM states obtained from the simulation with a99SB-UCB.

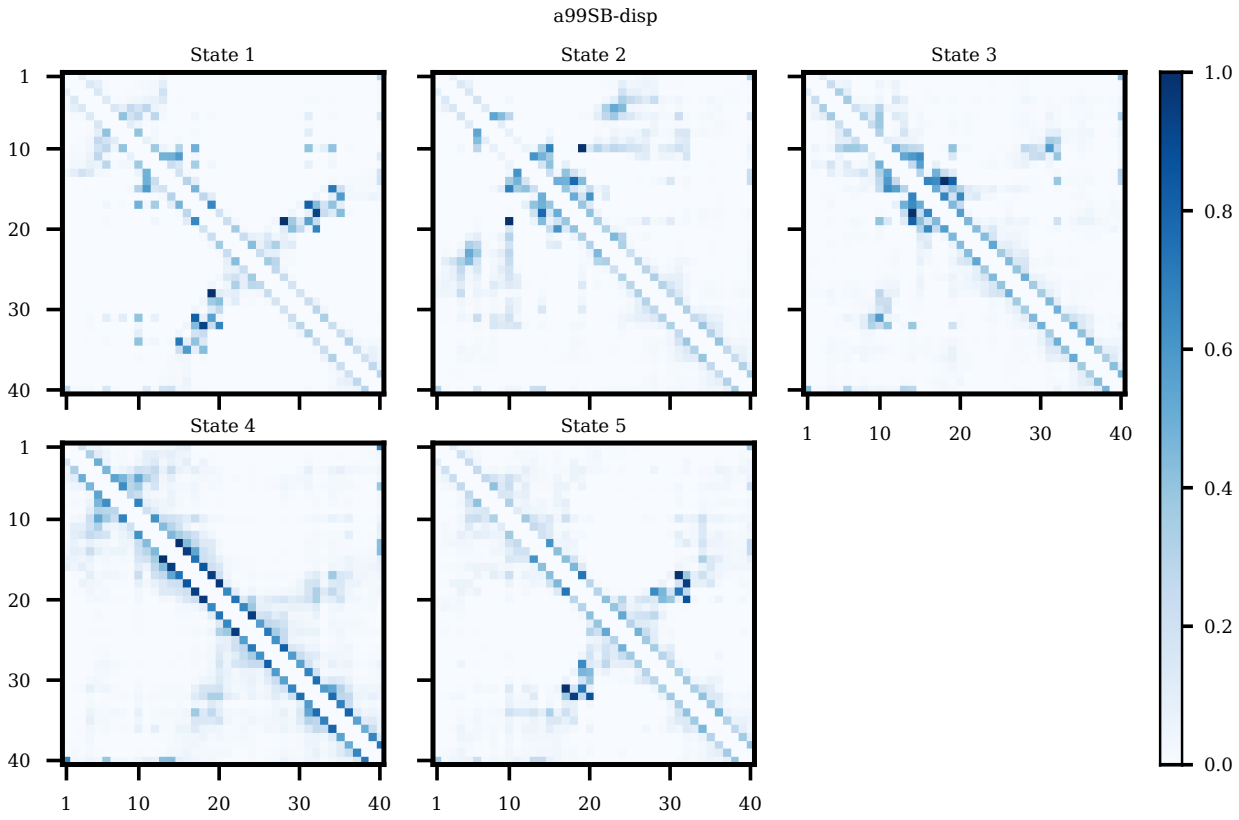

Figure S8: Normalized contacts for the coarse-grained MSM states obtained from the simulation with a99SB-disp.

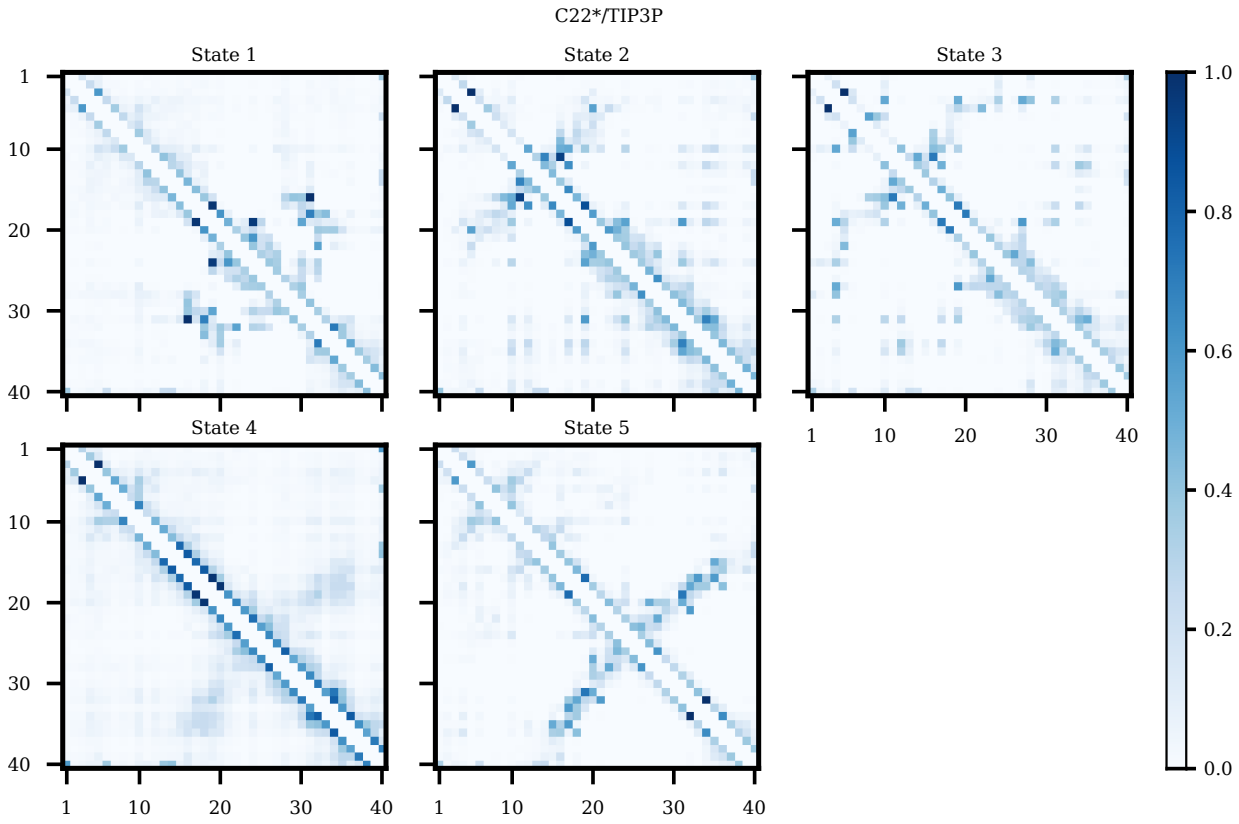

Figure S9: Normalized contacts for the coarse-grained MSM states obtained from the simulation with C22*/TIP3P.

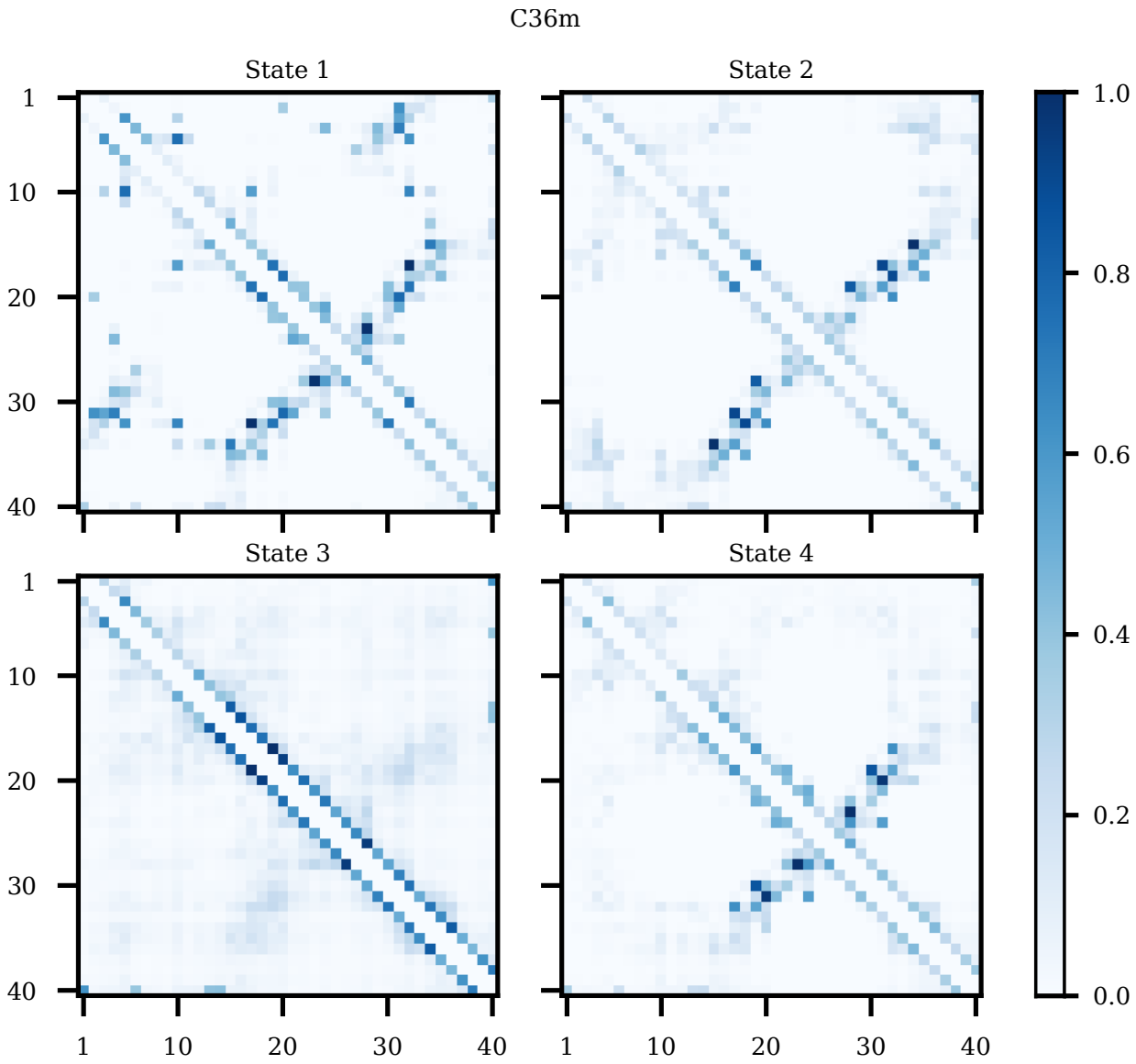

Figure S10: Normalized contacts for the coarse-grained MSM states obtained from the simulation with C36m.

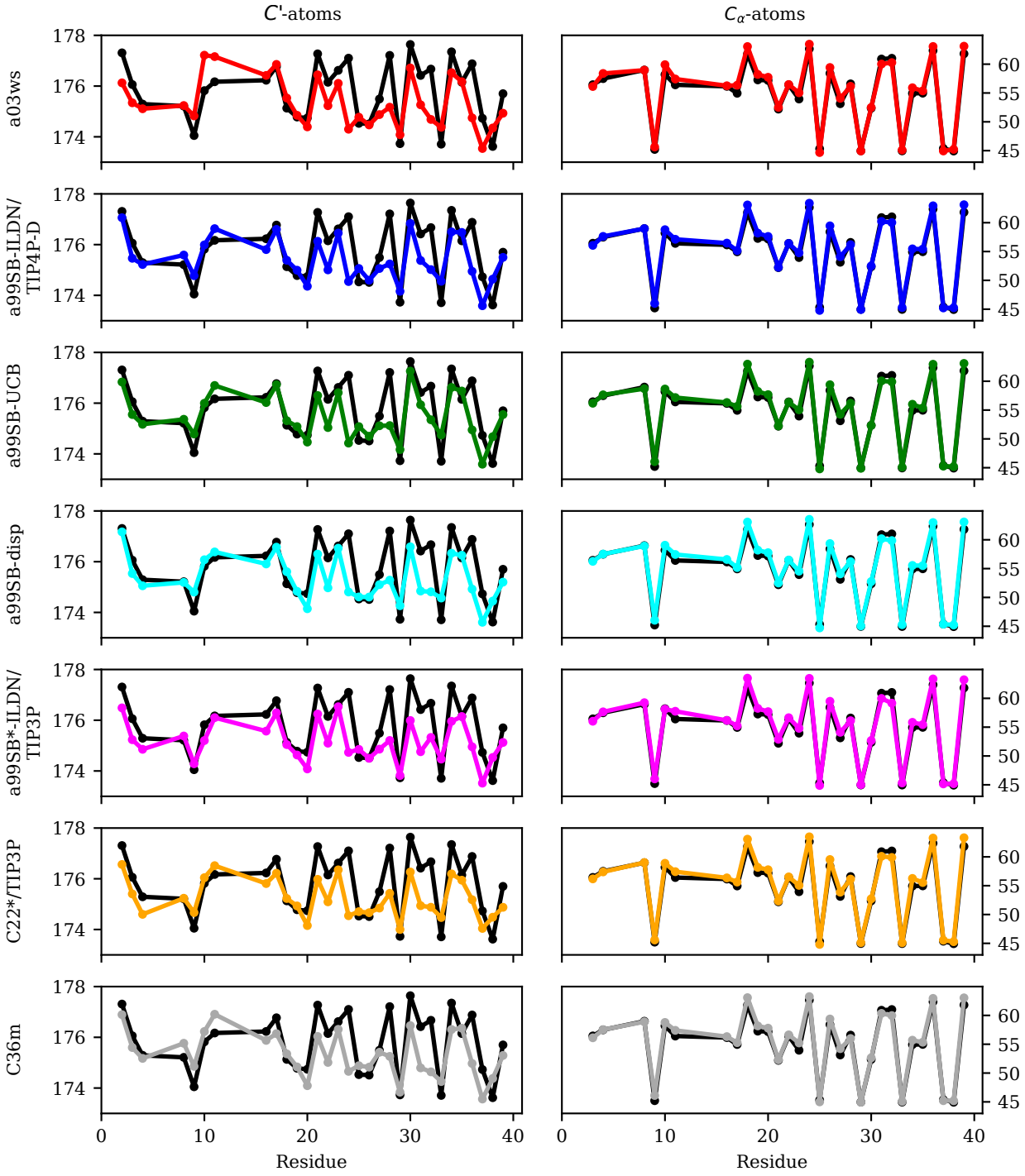

Figure S11: Experimental (black) and calculated (colored) NMR chemical shifts for the C′ atoms (left) and C_α atoms (right) for the different force fields (indicated on the left of each row).

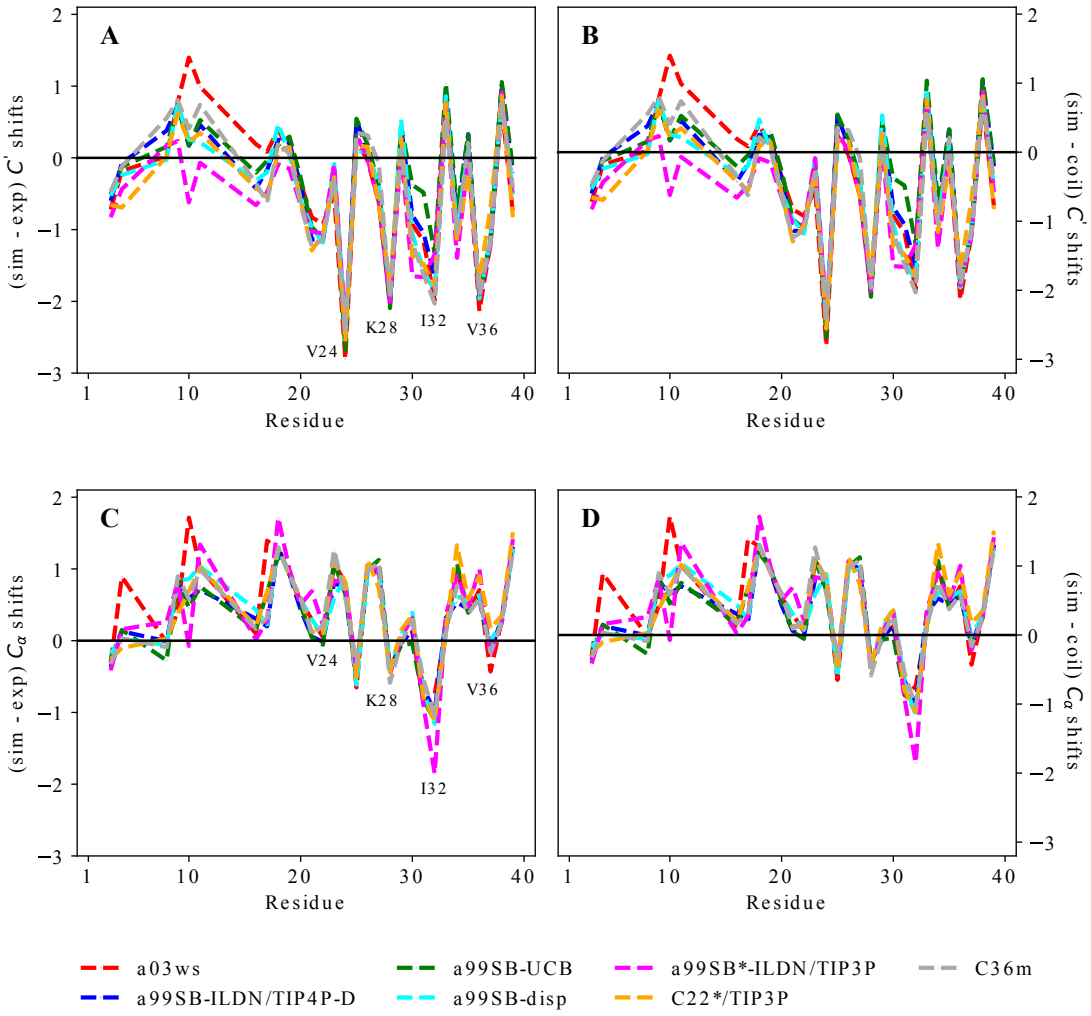

Figure S12: Difference between (A and C) the calculated and experimental and (B and D) the calculated and random coil NMR chemical shifts (i.e., the secondary chemical shifts) for the C′ and C_α atoms (top and bottom, respectively) of Aβ40 residues for the different force fields (see color key). The residues with the large deviations between simulation and experiment and which are discussed in detail in the text aare labeled in panels A and C.

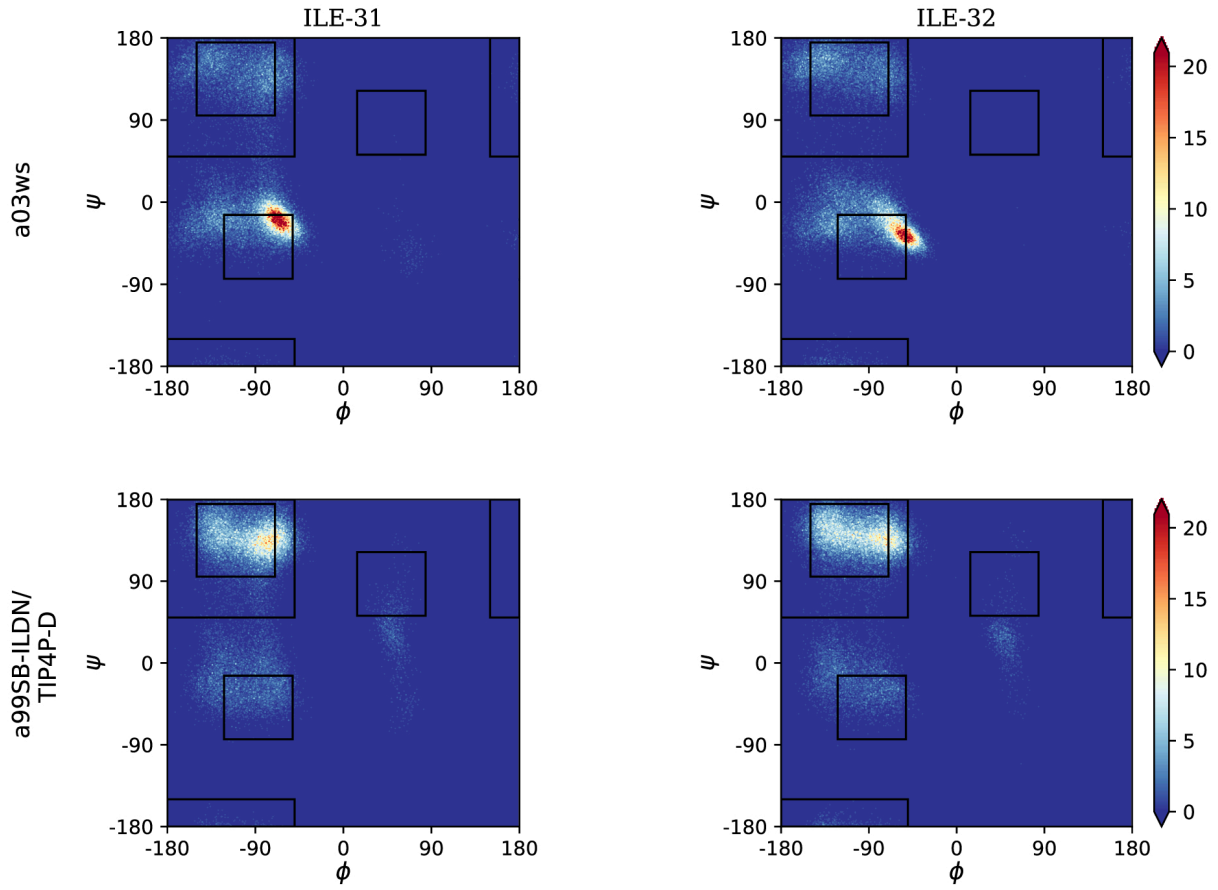

Figure S13: Ramachandran plots of I31 and I32 obtained from the simulation with a03ws (top) and a99SB-ILDN/TIP4P-D (bottom).

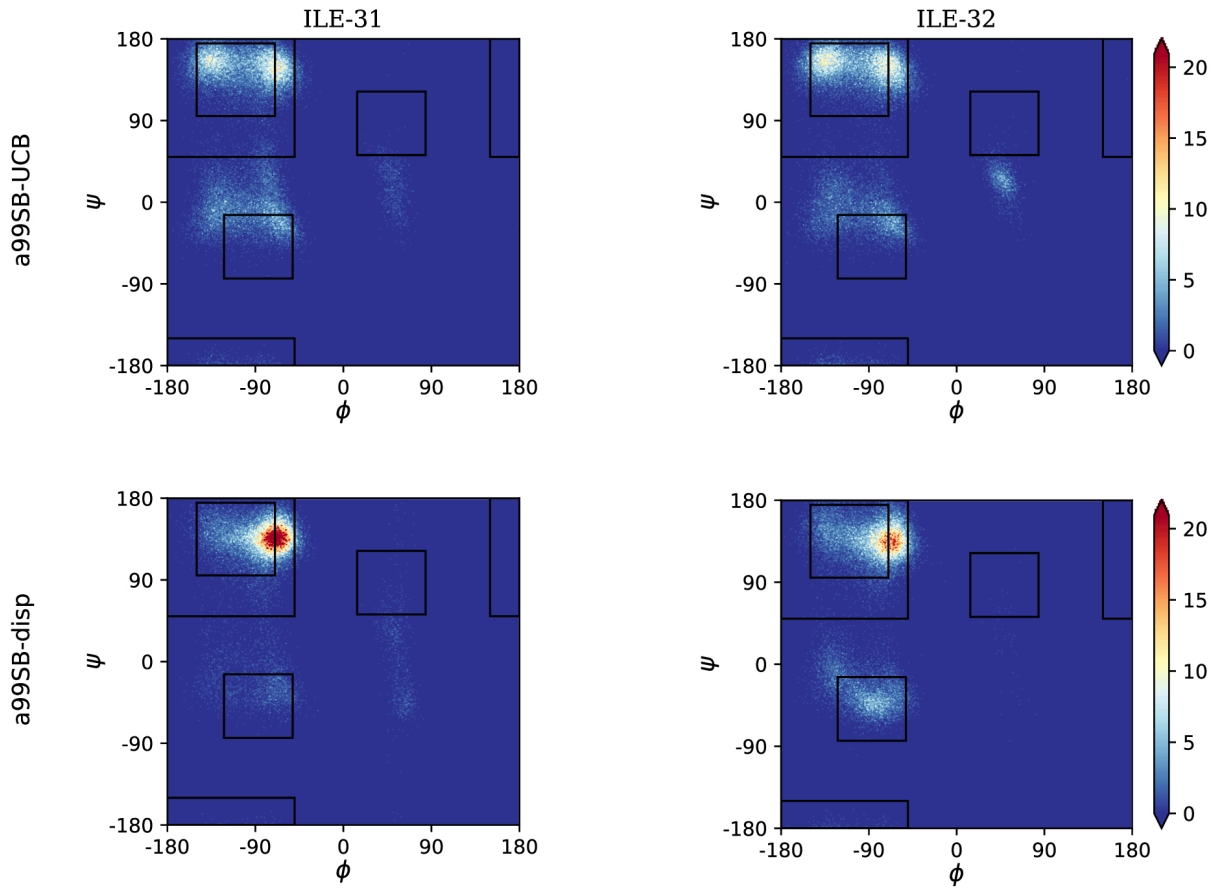

Figure S13: (cont.) Ramachandran plots of I31 and I32 obtained from the simulation with a99SB-UCB (top) and a99SB-disp (bottom).

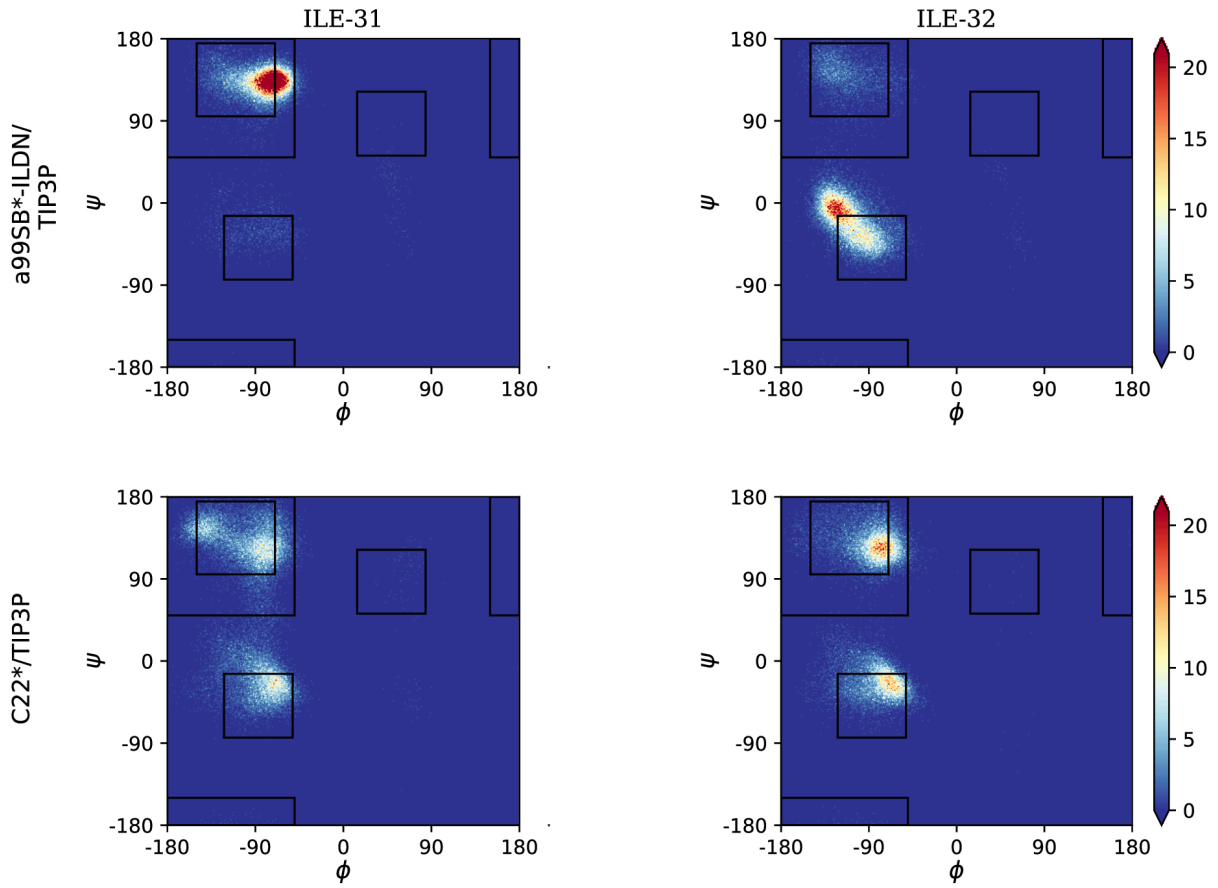

Figure S13: (cont.) Ramachandran plots of I31 and I32 obtained from the simulation with a99SB*-ILDN/TIP3P (top) and C22*/TIP3P (bottom).

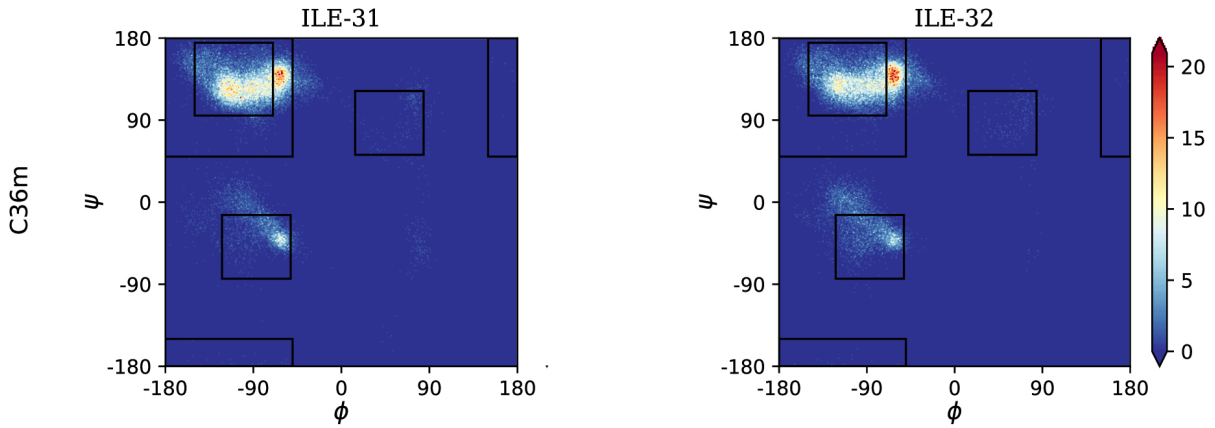

Figure S13: (cont.) Ramachandran plots of I31 and I32 obtained from the simulation with C36m.

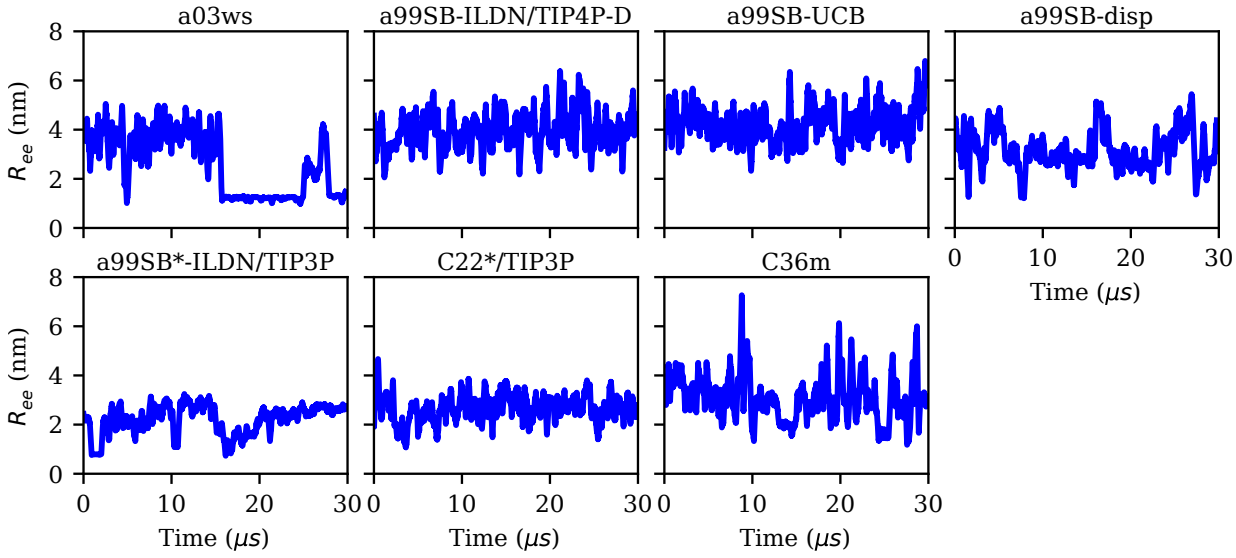

Figure S14: Evolution the end-to-end distance R_{ee} for the different force fields (labels on the top of the panels).

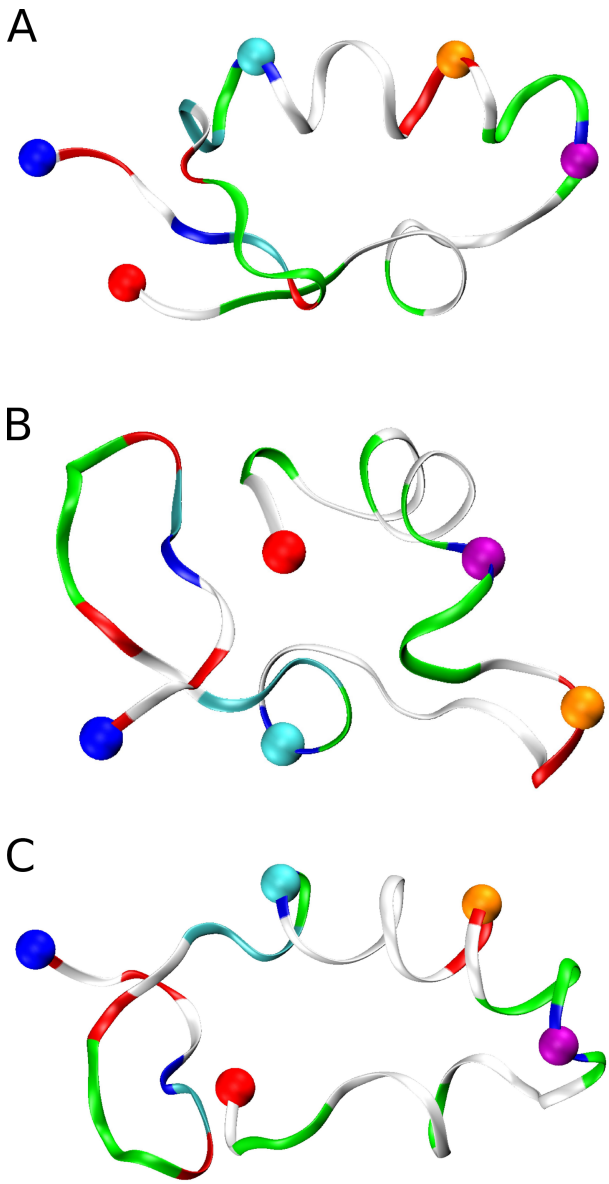

Figure S15: Compact Aβ40 structures sampled with a03ws between 16 and 25 µs. These conformations exhibit a high propenisty for helix formation in different parts along the sequence: (A) between residues K16 and K28 (as present in MSM state 1), (B) between residues G29 and M35 (as present in MSM states 3), (C) between residues K16 to K28 and G29 to M35 (as present in MSM state 2). Aβ40 is shown as band and colored according to amino acid residue type (basic: blue, acidic: red, histidine: cyan, polar: green, hydrophobic: white). Following residues are indicated by spheres: N-terminus (blue), K16 (cyan), D23 (orange), K28 (mauve), C-terminus (red).

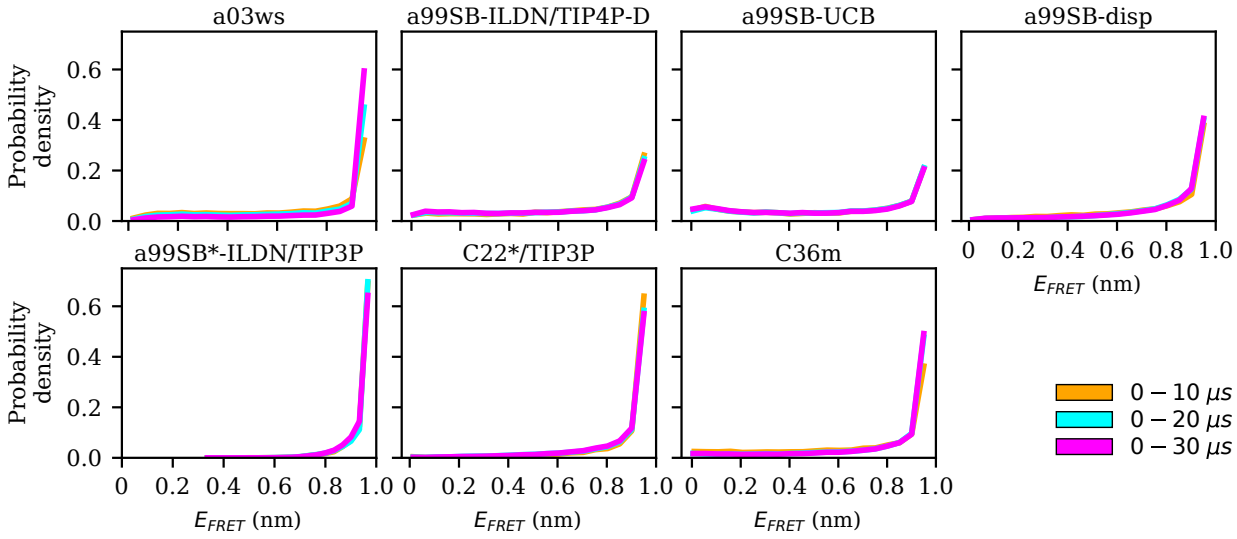

Figure S16: Distribution of the FRET efficiency E_{FRET} for increasing trajectory lengths (0–10 µs: yellow, 0–20 µs: cyan, 0–30 µs: magenta) for the different force fields (labels on the top of the panels).

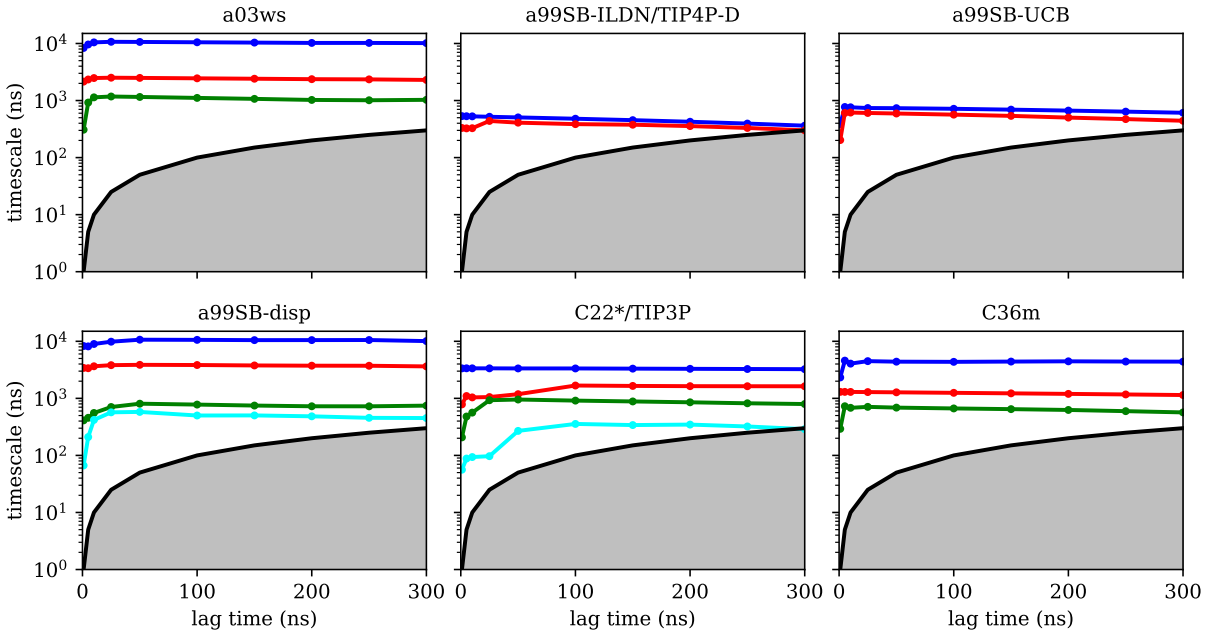

Figure S17: Implied time scales of the slowest processes (colored lines) obtained for different MSMs at different lag times (dots on colored lines) calculated from the MD trajectories using different force fields (labels on the top of the panels).

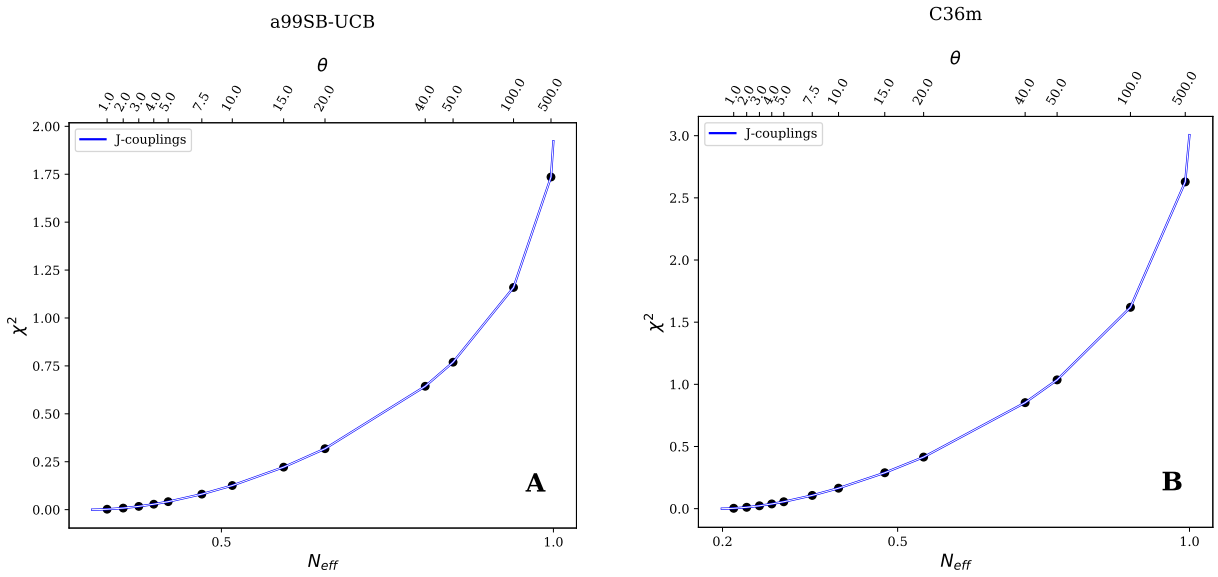

Figure S18: χ^2 versus the effective fraction of frames (N_{eff}) after reweighting the trajectory frames of the simulation with (A) a99SB-UCB and (B) C36m using the maximum entropy principle with different values of the parameter θ. In the Bayesian/Maximum Entropy (BME) procedure, one has to choose the parameter θ in such a way that the discrepancy between the experimental and simulation data is reduced (i.e., achieving a low χ^2 value) while minimally perturbing the initial set of weights of the trajectory, which corresponds to retaining a large effective fraction of the MD frames (N_{eff}). This is done by inspecting the χ^2 values vs. N_{eff}, which shows that when $\theta \to 0$, we have $\chi^2 \to 0$ but at the same time N_{eff} becomes very small ($N_{\text{eff}} \to 0$). On the other hand, for large θ, we see that χ^2 approaches the initial value (before reweighting) as the weights are least perturbed and as such N_{eff} is also close to 1. A trade-off between these two limits can be found by choosing different θ values, starting from a large one and reducing it until further decrease in θ does not result in a significant decrease in the χ^2 value. Following this, we chose $\theta = 10$ for both force fields.

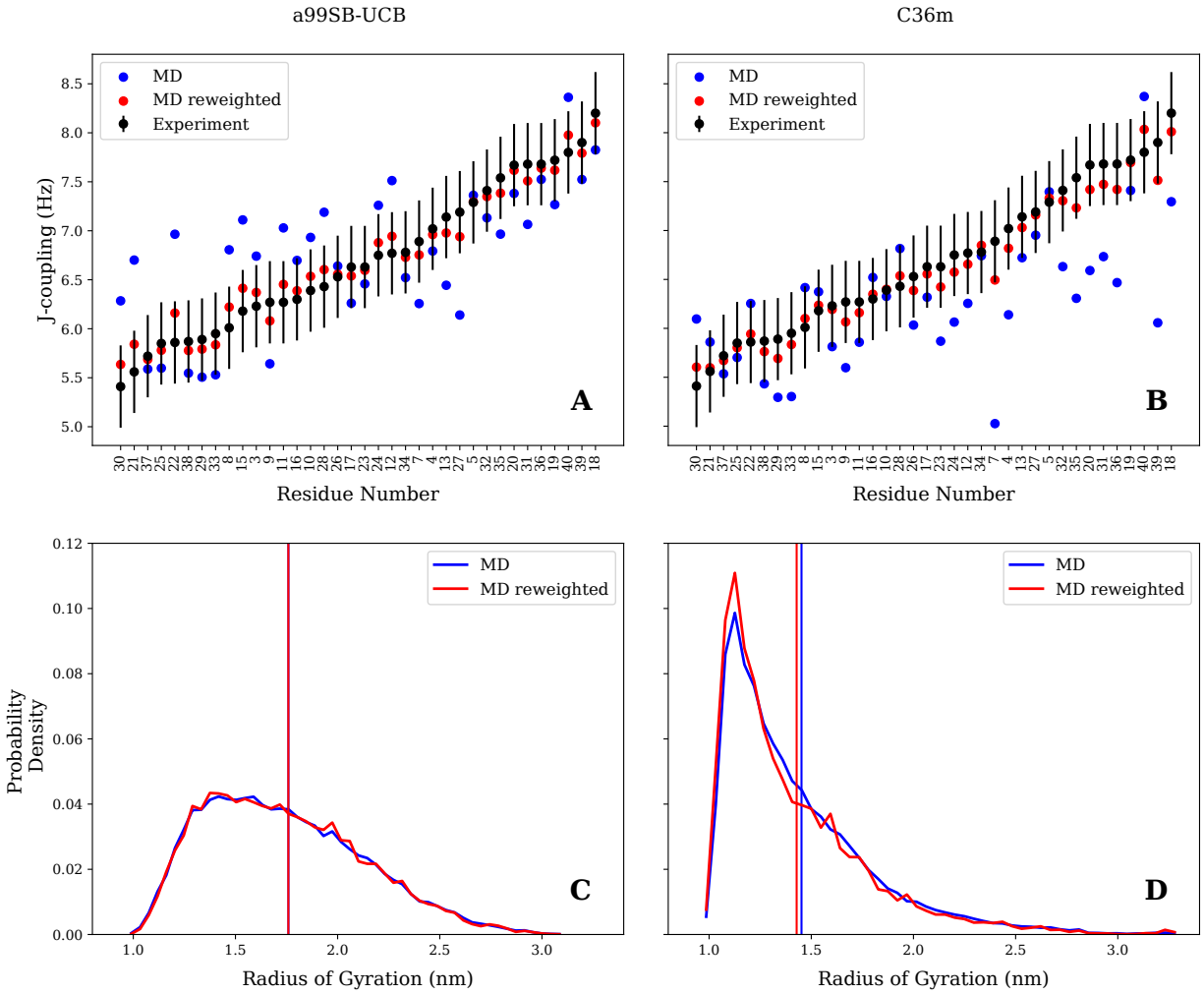

Figure S19: Reweighting of the trajectory frames using the maximum entropy principle to optimize the J-couplings obtained with the MD trajectory with a99SB-UCB (left) and C36m (right). (Top) The black dots indicate the experimental J-couplings for the individual Aβ40 residues (sorted in increasing J-coupling order), blue and red dots indicate the calculated J-couplings before and after, respectively, reweighting. (Bottom) Distribution of the radius of gyration before (blue) and after (red) reweighting the MD frames. The vertical lines indicate the corresponding R_{gyr} average.

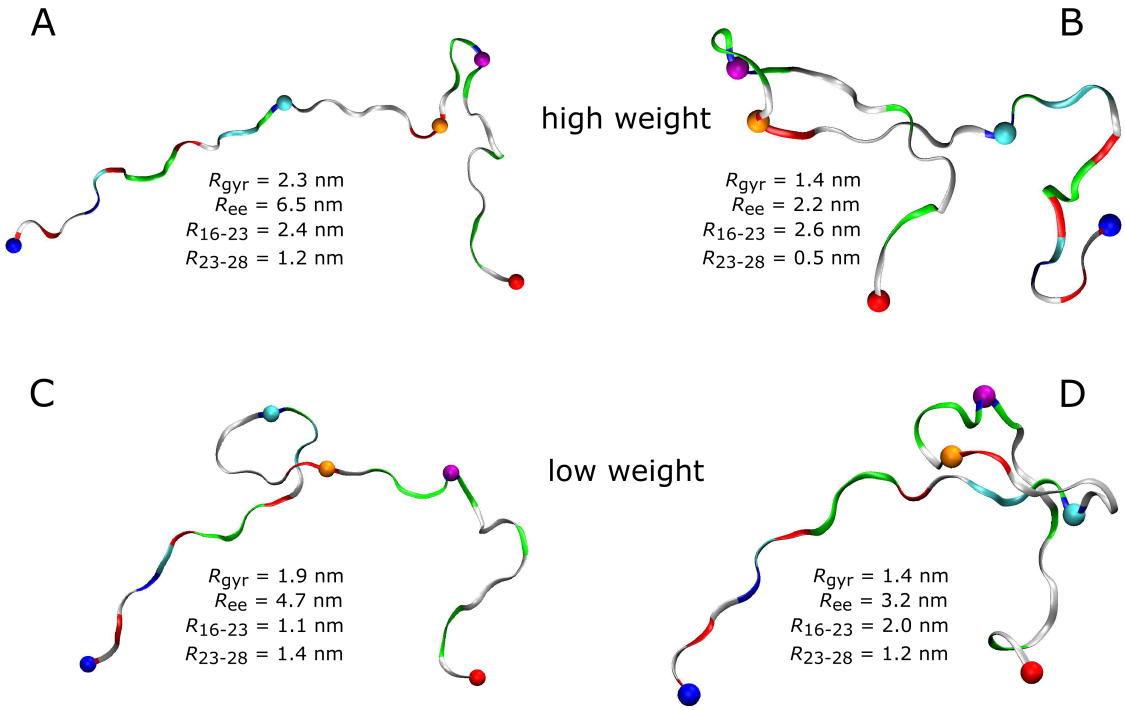

Figure S20: High-weight (A and B) and low-weight structures (C and D) determined by reweighting the C36m trajectory using the Bayesian/maximum entropy technique. Aβ40 is shown as band and colored according to amino acid residue type (basic: blue, acidic: red, histidine: cyan, polar: green, hydrophobic: white). Following residues are indicated by spheres: N-terminus (blue), K16 (cyan), D23 (orange), K28 (mauve), C-terminus (red). The structures were characterized in terms of R_{ee}, R_{gyr}, the K16–D23 distance (R_{16-23}), and the D23–K28 distance (R_{23-28}).

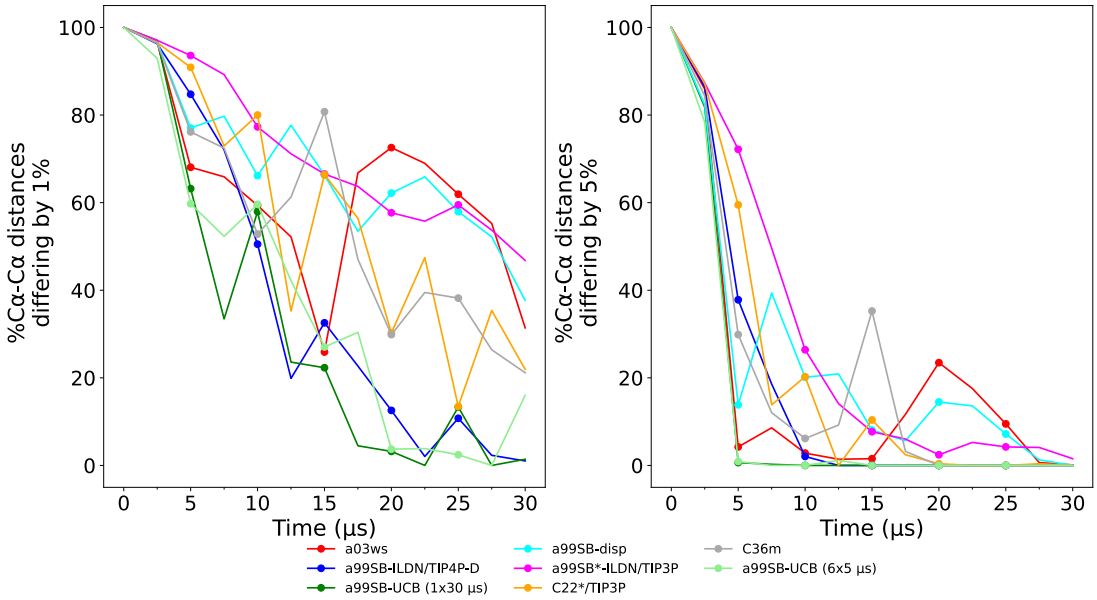

Figure S21: Measure of convergence based on time-averaged C_α–C_α distances. Averages were computed over increasing time window lengths in 2.5 μs increments. The percent of average distances that differ by more than 1% (left) or 5% (right) with respect to the preceding time interval are shown. If one applies the stricter 1% criterion, only a99SB-ILDN/TIP4P-D and a99SB-UCB yield converged C_α–C_α distances within 20–30 μs, which agrees to the conclusions drawn from the various other convergence tests. The second best set of force fields in terms of convergence of C_α–C_α distances is given by C22*/TIP3P and C36m, while the poorest convergence is observed for a03ws, a99SB-disp, and a99SB*-ILDN/TIP3P. If one applies the more generous 5% criterion, all force fields yield converged C_α–C_α distances, yet the order of the force fields with respect to convergence does not change. For instance, considerable motions are still observed between 20 and 25 μs for a03ws and a99SB-disp.

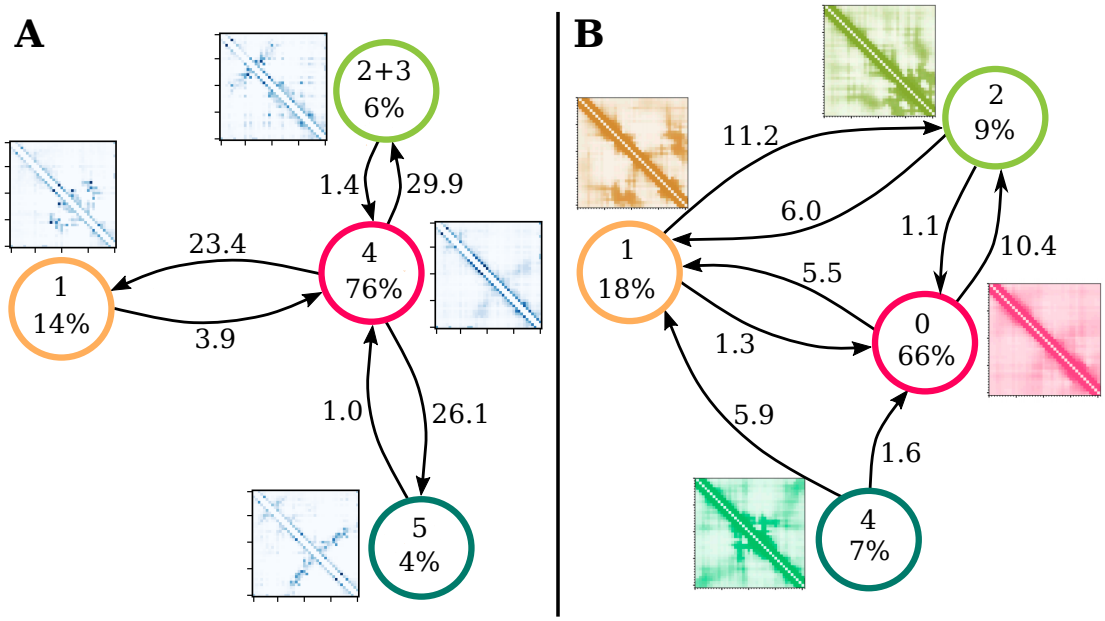

Figure S22: Comparison of (A) the MSM of Aβ40 determined in this work from the 30 µs MD simulation in conjunction with the C22*/TIP3P force field with (B) the MSM of Aβ42 derived from 5,119 trajectories between 9.75 and 90.5 ns in length with an aggregated simulation time of 315 µs, also based on C22*/TIP3P [1]. The numbers used for the states correspond to those used in this work and in [1], respectively. States 2 and 3 of the MSM for Aβ40 were combined into one state as they harbor very similar structures (see Figure S9). Both MSMs are dominated by the central state corresponding to extended Aβ structures, i.e., state 4 and 0 with a population of 76% and 66%, respectively. These two states have further in common that they tend to form a β-hairpin in the C-terminal half of the peptide. The second most populated state in either MSM is state 1, which is located left to the central state in the MSM, i.e., the transition from the central to this state involves a motion along TIC 1 in negative direction. The corresponding contact maps show that both states are characterized by interactions between residues of the C-terminal half. The states that emerge when going from the central states along TIC 2 in positive direction are the most different from each other when comparing (A) and (B). States 2 + 3 of Aβ40 are characterized by an N-terminal β-hairpin, while state 2 of Aβ42 is dominated by intense contacts limited to residues 30–42. State 5 of Aβ40 and state 4 of Aβ42, which are obtained by transitions from the central state along TIC 2 in negative direction, are rather similar again, with a β-hairpin in the C-terminal half being the dominating structural pattern. Differences between both MSMs are observed for some of the transitions and the accompanying MFPTs (in µs). Many of the MFPTs determined in this work are larger than those determined for Aβ42. This is probably caused by the different lag times used for constructing the MSMs: 100 ns in (A) and 12.5 ns in (B) [1]. For other proteins it was shown that lag times $\lesssim$ 100 ns tend to underestimate the MFPTs [2]. Other differences are that some of the transitions sampled for Aβ42, namely $4 \rightarrow 1$ and $1 \rightleftharpoons 2$ were not observed for Aβ40, while for Aβ42 the transition from the central state to state 4 was not sampled. These differences might arise from the fact that two different peptides are compared here, Aβ40 vs. Aβ42, or from the different sampling approaches, i.e., one long simulation vs. thousands of short simulations. Nonetheless, the overall conclusion is that both MSMs are rather similar to each other, suggesting that the 30 µs MD simulation of Aβ40 has reached convergence.

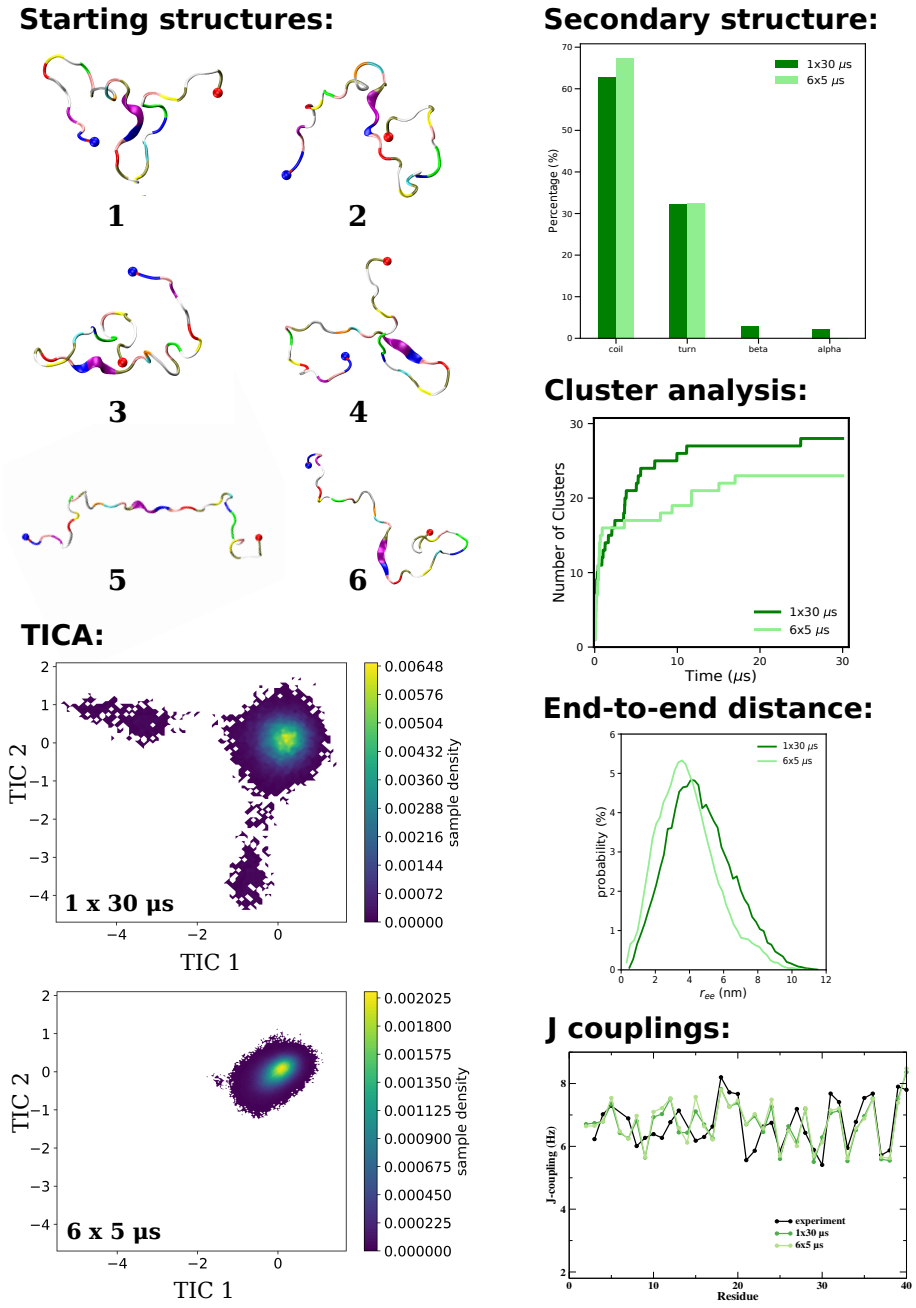

Figure S23: Results of 6×5 µs MD simulations using a99SB-UCB that were started from the structures **1** to **6**, which were collected at $t = 5, 10, \ldots, 30$ μs of the original 30 μs MD simulation using the same force field. The 6×5 µs trajectories were concatenated and the conformations project onto the TICs obtained from TICA applied to the original trajectory. One can see that only one state corresponding to extended Aβ40 conformations is sampled in these six MD runs, while the rare transitions to the states with β-sheet or α-helix did not take place. This can also be seen in the analysis of the secondary structure. Moreover, due to the missing transitions, fewer conformational clustes are identified. Nonetheless, the overall good agreement with experimental values still holds true. For instance, $\langle E_{\text{FRET}} \rangle = 0.58 \pm 0.11$, which is very close to the value of the 30 µs MD simulation and results from the similar R_{ee} distributions. Also the NMR results are nearly identical, as demonstrated here for the J-couplings ($\chi^2 = 2.28$).

References

1. T. Löhr, K. Kohlhoff, G. Heller, C. Camilloni and M. Vendruscolo, *Nature Comput. Sci.*, 2021, 1, 71–78.

2. E. Suárez, R. Wiewiora, C. Wehmeyer, F. Noé, J. Chodera and D. Zuckerman, *bioRxiv*, 2020, DOI:10.1101/2020.11.09.374496.