Dynamic Properties of Force Fields

F. Vitalini,[1,2] A.S.J.S. Mey,[1] F. Noé,[1,a)] and B.G. Keller[2,b)]

[1)] *Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany*

[2)] *Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Takustraße 3, D-14195 Berlin, Germany*[c)]

(Dated: 2 December 2014)

PACS numbers: Valid PACS appear here

Keywords: Suggested keywords

[a)]Electronic mail: frank.noe@fu-berlin.de

[b)]Electronic mail: bettina.keller@fu-berlin.de

[c)]Contributed equally to this work

I. METHODS

A. Capped Amino Acids (Ac-X-NHMe)

1. Simulation Details

We performed all-atom molecular dynamics simulations of acetyl-alanine-methylamide (Ac-A-NHMe) and acetyl-valine-methylamide (Ac-V-NHMe) in explicit water using the GROMACS 4.5.5 simulation package[1]. For both systems, simulations with five different potential energy function (force fields) were setup: AMBER ff99SB-ILDN[2], AMBER ff03[3], OPLS-AA/L[4], CHARMM27[5] and GROMOS43a1[6,7]. The water model was chosen to be in agreement with the one used for the validation of the force field, i.e. TIP3P[8] water model for AMBER ff99SB-ILDN, AMBER ff03, OPLS-AA/L and CHARMM27, and SPC[9] for GROMOS43a1. The simulations were performed in the NVT ensemble, where the temperature was restrained to 300 K using the V-Rescale thermostat[10].

Energy-minimized starting structures of Ac-A-NHMe and Ac-V-NHMe were solvated into (separate) cubic boxes with a minimum distance between solute and box walls of 1 nm, corresponding to a box length of about 2.76 nm^3 and 531 to 684 water molecules for Ac-A-NHMe, and 577 to 676 water molecules for Ac-V-NHMe. After an initial equilibration of

Force field	Ac-A-NHMe		Ac-V-NHMe		AVAVA		A_{10}	
	box	# H_2O	box	# H_2O	box	# H_2O	box	# H_2O
ff99SB-ILDN	20.12 nm^3	651	21.02 nm^3	676	50.70 nm^3	1684	233.45 nm^3	7647
ff03	19.68 nm^3	646	21.02 nm^3	672	50.70 nm^3	1684	233.45 nm^3	7647
OPLS-AA/L	21.02 nm^3	684	21.02 nm^3	672	54.23 nm^3	1703	233.45 nm^3	7647
CHARMM27	21.02 nm^3	680	21.02 nm^3	672	54.23 nm^3	1703	233.45 nm^3	7647
GROMOS43a1	16.77 nm^3	531	17.98 nm^3	577	50.65 nm^3	1703	233.45 nm^3	7647

TABLE I. Volume of the simulation box and number of water molecules per simulation box.

100 ps, 4 production runs of 1 μs each were performed for each simulation setup (exception: ff99SB-ILDN[2] force field, 20 production runs of 200 ns each), yielding a total simulation time of 4 μs for each of the ten simulation setups. Covalent bonds to hydrogen atoms were constrained using the LINCS algorithm[11] (lincs_iter = 1, lincs_order = 4), allowing

for an integration timestep of 2 fs. The leap-frog intergrator was used. Lennard-Jones interactions were cut off at 1 nm. Electrostatic interactions were treated by the Particle-Mesh Ewald (PME) algorithm[12] with a real space cutoff of 1 nm, a grid spacing of 0.15 nm, and an interpolation order of 4. Periodic boundary condition were applied in the x, y, and z-direction. The atom positions of the solute was written to file every 1 ps.

2. Markov State Models

Markov state models of the conformational dynamics were constructed on the space spanned by the ϕ and ψ backbone torsion angles of Ac-A-NHMe and Ac-V-NHMe, respectively. Time series of these dihedral angles were extracted from the simulated trajectories using the GROMACS command `g_rama`. The $\{\phi - \psi\}$-space was discretized using a regular grid with a binwidth of 10°, resulting in $36 \times 36 = 1296$ (equally sized) microstates states. The $\{\phi - \psi\}$-time series was projected onto this grid, yielding the corresponding microstate timeseries. The subsequent Markov state model analysis of this time series was performed using the EMMA software package[13] and consisted of the following steps. A list of the visited microstates which represented a connected set was generated using `mm_connectivity`. The list contained from 825 to 1059 microstates for the various setups of Ac-A-NHMe and from 644 to 1037 microstates for Ac-V-NHMe (Tab. II).

Force field	Ac-A-NHMe	Ac-V-NHMe	AVAVA	A_{10}
ff99SB-ILDN	825	721	494	2023
ff03	989	908	305	808
OPLS-AA/L	974	644	291	1265
CHARMM27	972	750	557	1325
GROMOS43a1	1059	1037	1231	2693
number of microstates	1296	1296	1536	6561

TABLE II. Size of the connected sets and total number of microstates of the MSM discretizations.

The actual transition matrices were estimated on this reduced set of microstates (option `-restrictToStates`), using a sliding window algorithm (option `-slidingwindow`) and

enforcing reversibility (option `-reversible`). Implied timescales

$$t_i(\tau) = \frac{-\tau}{\ln(\lambda_i(\tau))} \tag{1}$$

were estimated for the eigenvalues λ_2 and λ_3, with lag times ranging from $\tau = 1$ ps to $\tau = 100$ ps using the command `mm_timescales` The transition matrices used for further analysis (eigenvectors and eigenvectors) were estimated using the command `mm_estimate` with a lag time of $\tau = 50$ ps. Only for Ac-A-NHMe simulated with the GROMOS43a1 force field, the lag time had to be set to $\tau = 10$ ps, because at $\tau = 50$ ps the third eigenvalue had already decayed to (a numerical) zero. These transition matrices were further analyzed using `mm_transitionmatrixAnalysis`, i.e. the stationary density, the first 3 eigenvalues, left eigenvectors and right eigenvectors were written to file.

The eigenvectors were projected onto the original set of $36 \times 36 = 1296$ microstates, assigning a value of "0" to the microstates not included in the reduced set of microstates. The EMMA software package[13] yields eigenvectors which are normalized with respect to the Euclidean scalar product; they were renormalized to ensure orthonormality with respect to the following scalar products

$$\begin{aligned} \left\langle \mathbf{l}_i, \Pi^{-1}\mathbf{l}_j \right\rangle &= \delta_{ij} \\ \langle \mathbf{r}_i, \Pi\mathbf{r}_j \rangle &= \delta_{ij}\,, \end{aligned} \tag{2}$$

where δ_{ij} is the Kronecker delta, and $\Pi = \mathrm{diag}(\pi_1, \pi_2, ...\pi_n)$ is a diagonal matrix which has the stationary distribution π along its diagonal. Consequently, $\Pi^{-1} = \mathrm{diag}(1/\pi_1, 1/\pi_2, ...1/\pi_n)$. The post-processing of eigenvalues and eigenvectors was implemented in Python and MATLAB.

B. Pentapeptide AVAVA

1. Simulation Details

We performed all-atom molecular dynamics simulations of a penta-peptide with the sequence AVAVA in explicit water using the GROMACS 4.5.5 simulation package[1]. The protonation state of the end groups was set -NH_2 for the N-terminus and to -COOH for the C-terminus. Again, five different simulation setups were used which differed in the potential energy function used to model the forces in the simulation (i.e. the force field): AMBER

ff99SB-ILDN[2] , AMBER ff03[3], OPLS-AA/L[4], CHARMM27[5] and GROMOS43a1[6,7]. The water model was chosen to be in agreement with the water model used for the validation of the force field, i.e. TIP3P[8] water model for AMBER ff99SB-ILDN, AMBER ff03, OPLS-AA/L and CHARMM27, and SPC[9] for GROMOS43a1. The simulations were performed in the NVT ensemble at a temperature of 300 K. The remainder of the simulation setup was analogous to the one described in section I A 1 and differed only in the following parameters. Extended instead of energy minimized structures were solvated, to assure the independence from its periodic copies of the system. The size of the resulting simulation boxes and the number of water molecules are reported in Tab. I. For each of the five setups, two production runs of 2 μs each were performed, yielding to 4 μs simulation time for each setup (exception GROMOS43a1, 4 independent runs of 1 μs each).

2. Markov Model Construction

Markov state models of the conformational dynamics were constructed on the space spanned by the five pairs of ϕ- and ψ-backbone dihedral angles. Each ϕ-ψ plane was discretized into 6 (residue 1) or 4 (residue 2 to 5) bins resulting in a total number of $6 \times 4 \times 4 \times 4 \times 4 = 1536$ microstates. In Fig. 1, the discretization is plotted on top of the ϕ-ψ-distribution of each residue, showing that the maxima of this distribution fall into the the defined bins. For residue 2 to 4, the bins represent the following backbone conformations: bin 0 = β-sheet, bin 1 = α-helix, bin 2 and 3 = left-handed α-helix. The ϕ-ψ-distribution of residue 1 is structured by the rotation of the amino group around the ϕ-dihedral angle, which exhibits three stable conformations. To allow direct comparison, the same discretization was used for all force fields. Time series of these dihedral angles were extracted from the simulated trajectories using the python library MD_Analysis[?] and projected onto the grid of microstates using a Python script. The subsequent construction of Markov state models and the analysis of the eigenvectors and eigenvalues were analogous to section I A 2. The number microstates in the largest connected set of each simulation are shown in Tab. II. The eigenvalues and associated eigenvectors were evaluated at lag time $\tau = 2$ ns, except for GROMOS43a1 where a shorter lag time of $\tau = 0.5$ ns had to be used.

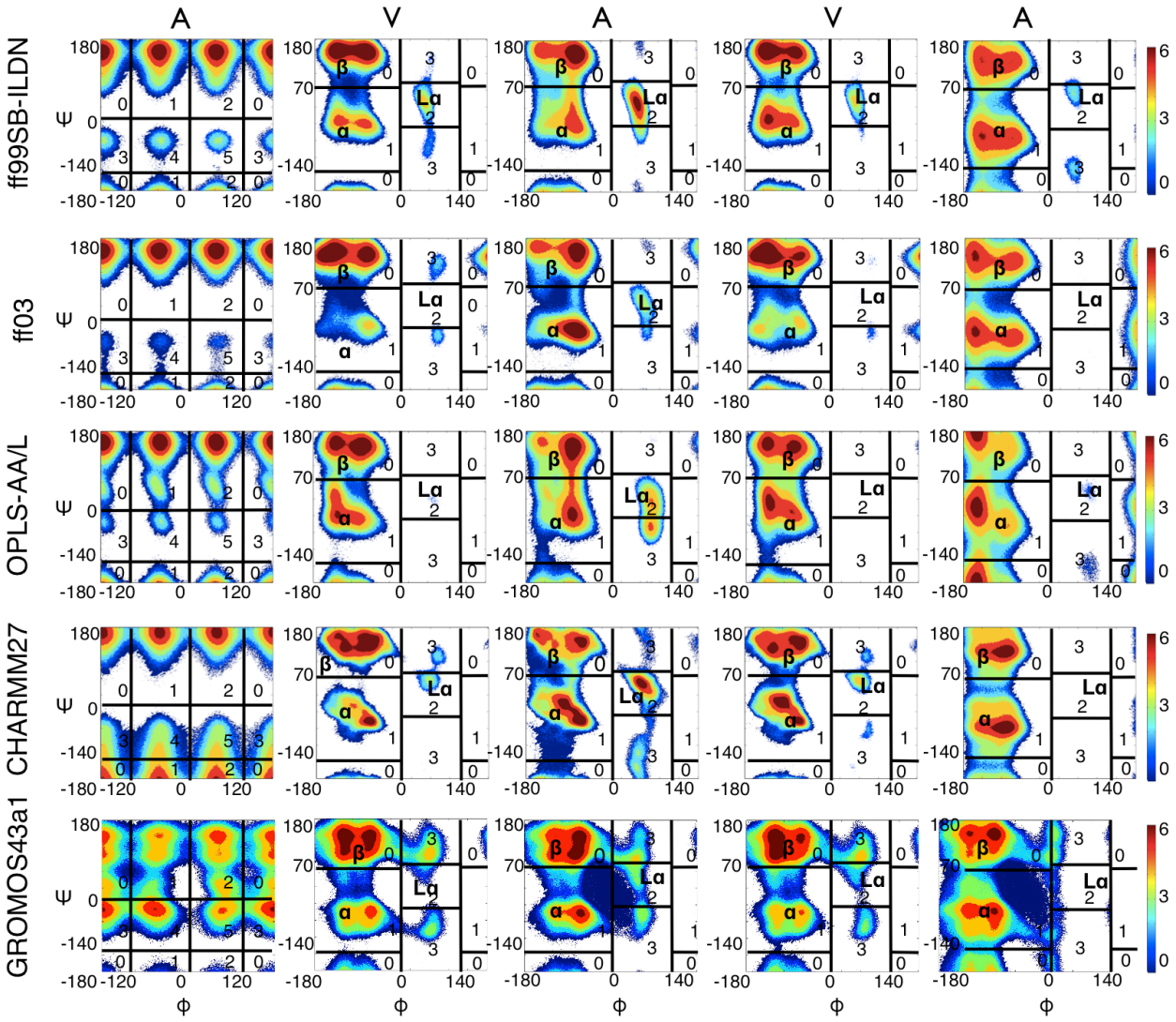

FIG. 1. Discretization of the conformational space of AVAVA. For each residue, the logarithmic counts of the distribution of the ϕ-ψ-dihedral angle space is discretized into four states according to secondary structure elements. Exception is the first alanine residue; being a terminal residue its dynamics is different. For this residue a six-states grid is applied. The same gird is able to capture the relevant minima in all force fields.

C. Deca-alanine

1. Simulation Details

We performed all-atom molecular dynamics simulations of a deca-peptide with the sequence A_{10} (deca-alanine) in explicit water using the GROMACS 4.5.5 simulation package[1].

Five different simulation setups were used, as before, each setup differing only in the choice of force fields, which were AMBER ff99SB-ILDN[2] , AMBER ff03[3], OPLS-AA/L[4], CHARMM27[5] and GROMOS43a1[6,7]. The water model was chosen to be in agreement with the water model used for the validation of the force field, i.e. TIP3P[8] water model for AMBER ff99SB-ILDN, AMBER ff03, OPLS-AA/L, and CHARMM27, and SPC[9] for GROMOS43a1. The simulations were performed in the NVT ensemble at a temperature of 300 K. The remainder of the simulation setup was analogous to the one described in section I A 1, where small deviations are discussed in the following. Extended or 'unfolded' structures were solvated in a box consisting of 7647 TIP3P water molecules for the two AMBER force fields, OPLS-AA/L, and CHARMM27. For the GROMOS43a1 setup 7647 SPC water molecules were used. The box dimensions were chosen such that self interaction across the periodic image did not occur, resulting in a box of size: 233.45 nm^3 for all force fields. From the same extended initial state a short equilibration simulation of 100 ps was carried out in the NVT ensemble from which the production runs for each force field were initiated. A set of eight 500 ns long trajectories for each force field was generated, resulting in a total simulation time of 20 μs for the deca-alanine (4 μs for each setup).

2. Markov Model Construction

Similar to the Markov model construction of the AVAVA in sec I B 2, for the deca-alanine, we chose a discretization in the Ramachandran space (Fig.**?? a,b** in the main part of the manuscript). Disregarding the two charged end groups deca-alanine has eight dihedral ϕ-ψ angle planes, each of which has been discretized into 3 bins. This resulted in a total number of $3^8 = 6561$ microstates. The three bins in each plane correspond to the following backbone conformations: bin 0 = α-helix, bin 1 = β-sheet, and bin 2 = L_α (Fig. **?? b** in the main part of the manuscript). The same discrete set of states for all force fields appears to be a valid choice, as the maxima of the joint dihedral distribution fall into the defined states for each of the force fields. Timeseries of these dihedral angles were extracted from the simulated trajectories using the GROMACS command `g_rama`. Projecting the eight ϕ-ψ dihedral angle pairs onto the grid yielded a ternary representation of conformation, e.g. 01121101. Converting this number to base ten yielded the microstate number. The analysis of the resulting microstate trajectory was analogous to section I A 2. The number microstates in

the largest connected set of each simulation are shown in Tab. II. The eigenvalues and eigenvectors were evaluated at lag time τ= 2 ns.

D. Error Estimation

The statistical uncertainty in the implied timescales (Figs. 2B, 3B, 4B, and 5B) was estimated using a bootstraping apporach. Boostrapping, as all resampling methods, requires subsampling of the existing data in order to extract statistical errors. In this case for each simulated system, out of the 4 μs data 8 continuous trajectory snapshots each of length 500 ns were drawn, in order to yield a total simulation time of 4 μs; exception is AMBER ff99SB-ILDN for the capped amino acids, where the maximum independent run was 200 ns; consequently 20 continuous trajectory snapshots of length 200 ns were used in this case, leading again to a total simulation time of 4 μs. This means that some trajectories snippets may be used multiple times, as a drawing with replacement occurs. The new set of trajectories contains the same number of data points as the original data set, albeit with a slightly different sampling of the transitions. From the new data set, transition matrices and implied timescales were calculated at range of lag times τ. The procedure was repeated 50 times to obtain a sample of implied time scale estimates at each lag time. The mean and standard deviation of the sample were estimated and are plotted in Figs. 2**B**, 3**B**, 4**B**, and 5**B**.

II. SUPPLEMENTARY RESULTS

A. Convergence Checks and Validation

Figs. 2, 3, 4, and 5 show test for the convergence of the MD simulation and for the validity of the Markov state models. The first plot in panel **A** shows the number of visited microstates as a function of total simulation time T, and should be compared to the second plot in panel **A**, which shows the the contribution of the visited microstates to the equilibrium distribution. Although the number of visited microstates increases throughout the entire simulation time for all simulation setups (with the possible exception of the OPLS-AA/L simulation of AVAVA), after 10 to 100 ns (roughly 1% of the total simulation time), the contribution of the newly "discovered" microstates to the equilibrium distribution is negligible. We can therefore conclude that the equilibrium distribution is sampled to convergence.

Remarkably, the GROMOS43a1 force field samples appreciably more microstates than all other force fields in this study. To verify that this observation is not caused by a single trajectory, which happens to sample a large number of otherwise rarely visited states, we estimated the expectation value and the standard deviation of the number of visited microstates as a function of simulation time. The results are shown in the third plot of Fig. 4A and 5**A**. GROMOS43a1 initially samples a the microstates at a much larger rate than the other four force fields, which is in agreement with the broader minima of GROMOS43a1 in the potential energy surface of the ϕ-ψ space. After roughly 0.5 ns the sampling rate, however, levels off to a rate comparable to that of the samplings rates of the other force fields in this study.

Panel **B** of Fig. 2, 3, 4, and 5 shows the implied time scale corresponding to the second and third eigenvalue as a function of the lag time (green and blue lines). The statistical uncertainty (estimated as the standard deviation of bootstrap samples) is shown as shaded areas. The line in which the implied time scale is equal to the lag time is shown as a black line in each plot. Below this line, the implied time scale estimates are increasingly dominated by numerical inaccuracy. Since in a true Markov process, the implied time scales are independent of the lag time (corresponding to a horizontal line in the implied time scale plot), these plots were used to find a range of lag times for which the dynamics of the

molecule is approximately markovian. With the exception of the ff03 simulation of Ac-V-NHMe and the GROMOS43a1 simulation of A_{10}, all implied time scale plots show extended plateau regions, verifying that the dynamics can be modeled as a Markov process in the chosen set of microstates. We can speculate for Ac-V-NHMe that the dynamics of the side-chain, not considered in this study, has a non-negligible contribution to the slowest kinetic process.

Panel **C** of Figs. 2, 3, 4, and 5 shows the convergence of the eigenvectors as a function of lag time τ. As a measure we used the Euclidean distance of the eigenvector estimated at some lag time τ to the corresponding eigenvector of the transition matrix with lag time τ' which is used to illustrate the properties of the force field in the main part of the manuscript. These plots should be compared to the matrices in Panel **D**, which show the Euclidean distance between the eigenvectors. Note that these distances are at least one order of magnitude larger than the drift of the eigenvectors with increasing lag time.

B. Comparison of the Kinetic Processes in Ac-A-NHMe and Ac-V-NHMe Across The Force Fields

Fig. 6 - 11 show difference plots of the equilibrium distribution and the two slowest kinetic processes in Ac-A-NHMe and Ac-V-NHMe. They highlight systematic differences in the sampling of the Ramachandran plane between the force fields.

The slowest kinetic processes of AVAVA and A_{10} are shown in Fig. 12 and 13. The conformational space is too complex to visualize the entire process. We therefore selected representative microstates: the three most populated microstates to which the corresponding eigenvectors assigns positive values, and the three most populated microstates to which the eigenvector assigns negative value. The process is then interpreted as the kinetic exchange between the conformations associated to these two sets of microstates. For both systems, the slowest processes differ considerably across force fields.

REFERENCES

[1]D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: fast, flexible, and free." J. Comput. Chem. **26**, 1701–18 (2005).

[2]K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field." Proteins **78**, 1950–8 (2010).

[3]Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. E. I. Zhang, R. Yang, P. Cieplak, R. A. Y. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, "A Point-Charge Force Field for Molecular Mechanics Quantum Mechanical Calculations," J. Comput. Chem. **24**, 1999–2012 (2003).

[4]G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Comparison with Accurate Quantum Chemical Calculations on Peptides ," J. Phys. Chem. B **2**, 6474–6487 (2001).

[5]a. D. MacKerell, N. Banavali, and N. Foloppe, "Development and current status of the CHARMM force field for nucleic acids." Biopolymers **56**, 257–65 (2001).

[6]X. Daura, A. E. Mark, and W. F. V. A. N. Gunsteren, "Parametrization of Aliphatic CH n United Atoms of GROMOS96 Force Field," J. Comp. Chem. **19**, 535–547 (1998).

[7]W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren, "The GROMOS Biomolecular Simulation Program Package," J. Phys. Chem. A **103**, 3596–3607 (1999).

[8]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926 (1983).

[9]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The Missing Term in Effective Pair Potentials," J. Phys. Chem. **91**, 6269–6271 (1987).

[10]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling." J. Chem. Phys. **126**, 014101 (2007).

[11]B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," J. Comput. Chem. **18**, 1463–1472 (1997).

[12]T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald : An N -log (N) method for Ewald sums in large systems," J. Phys. Chem. **98**, 10089–10092 (1993).

[13]M. Senne, B. Trendelkamp-Schroer, A. S. J. S. Mey, C. Schütte, and F. Noé, "EMMA : A Software Package for Markov Model Building and Analysis," J. Chem. Theory Comput. **8**, 2223–2238 (2012).

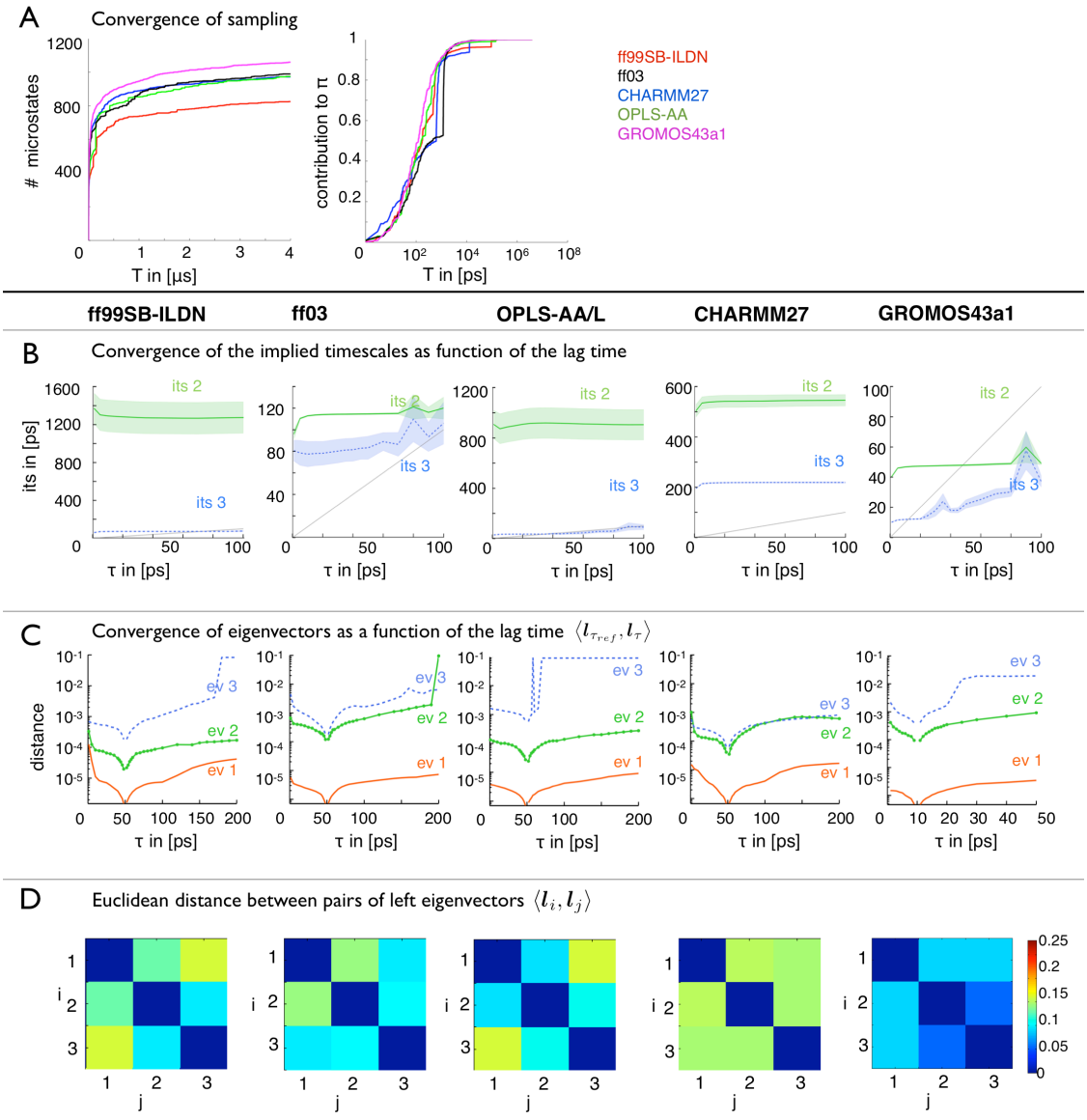

FIG. 2. Convergence checks for the simulation and Markov state models of Ac-A-NHMe. **A:** Convergence of the MD sampling with respect to the total simulation time T. **B:** Convergence of the implied timescales of the MSM. **C:** Convergence of the left eigenvectors of the MSM. **D:** Euclidean distance of pairs of the left eigenvectors of the MSM at lag time $\tau = 50$ ps (exception GROMOS43a1, τ= 10 ps).

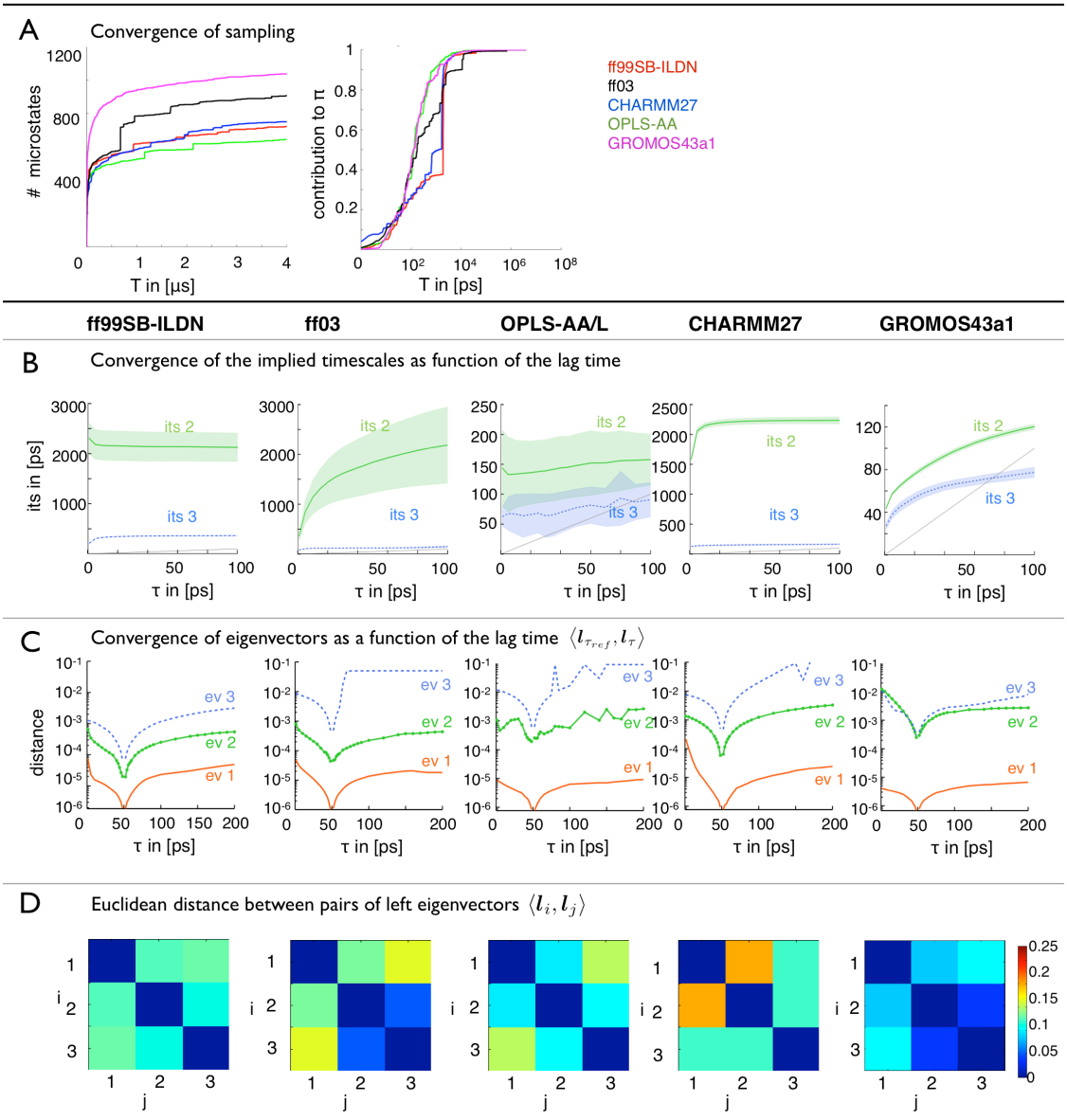

FIG. 3. Convergence checks for the simulation and Markov state models of Ac-V-NHMe. **A:** Convergence of the MD sampling with respect to the total simulation time T. **B:** Convergence of the implied timescales of the MSM. **C:** Convergence of the left eigenvectors of the MSM. **D:** Euclidean distance of pairs of the left eigenvectors of the MSM at lag time $\tau = 50$ ps (exception GROMOS43a1, τ= 10 ps).

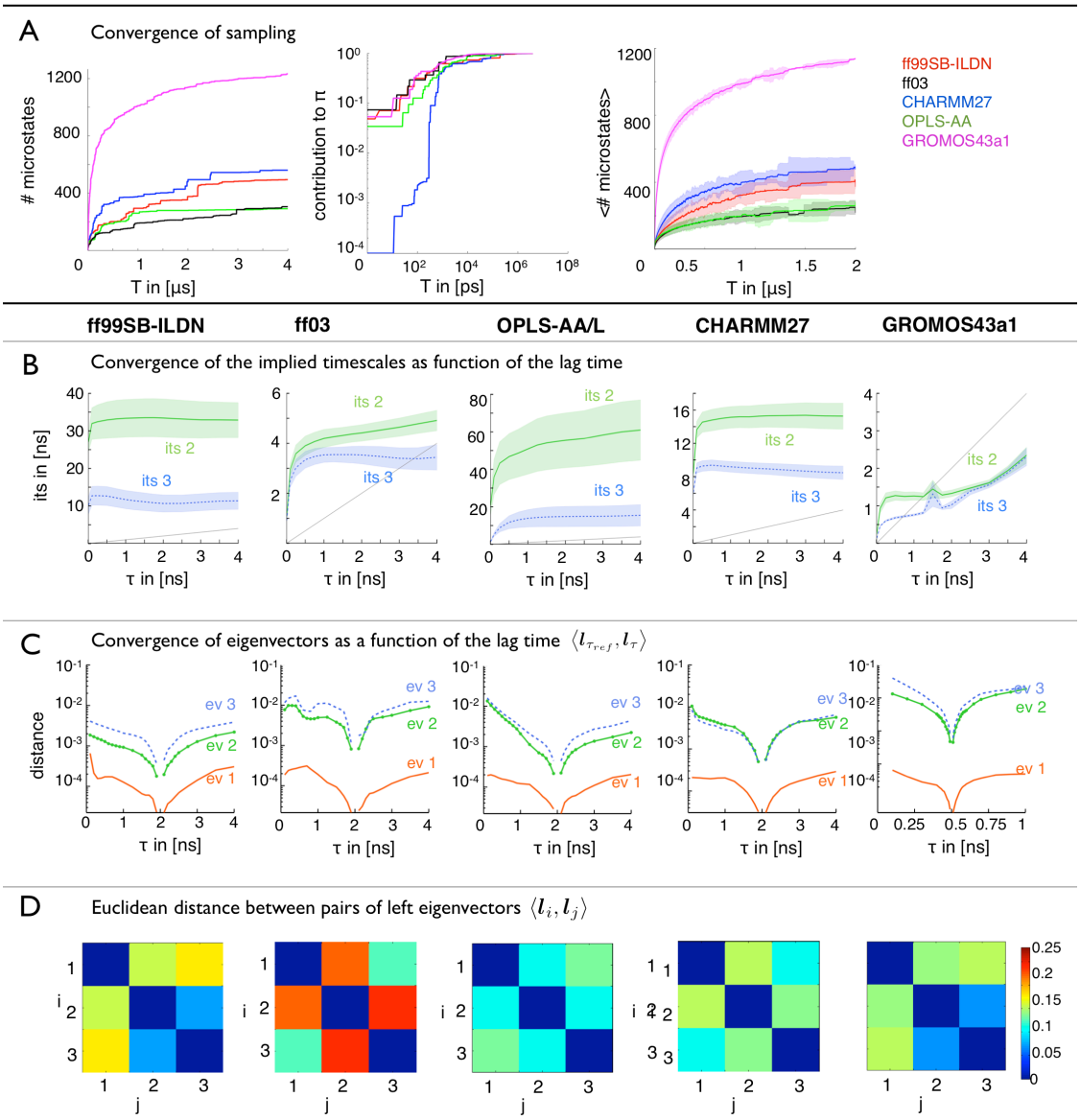

FIG. 4. Convergence checks for the simulation and Markov state models of the penta-peptide AVAVA. **A:**Convergence of the MD sampling with respect to the total simulation time T. **B:** Convergence of the implied timescales of the MSM. **C:** Convergence of the left eigenvectors of the MSM. **D:** Euclidean distance of pairs of the left eigenvectors of the MSM at lag time $\tau = 2$ ns.(exception GROMOS43a1, τ= 0.5 ns)

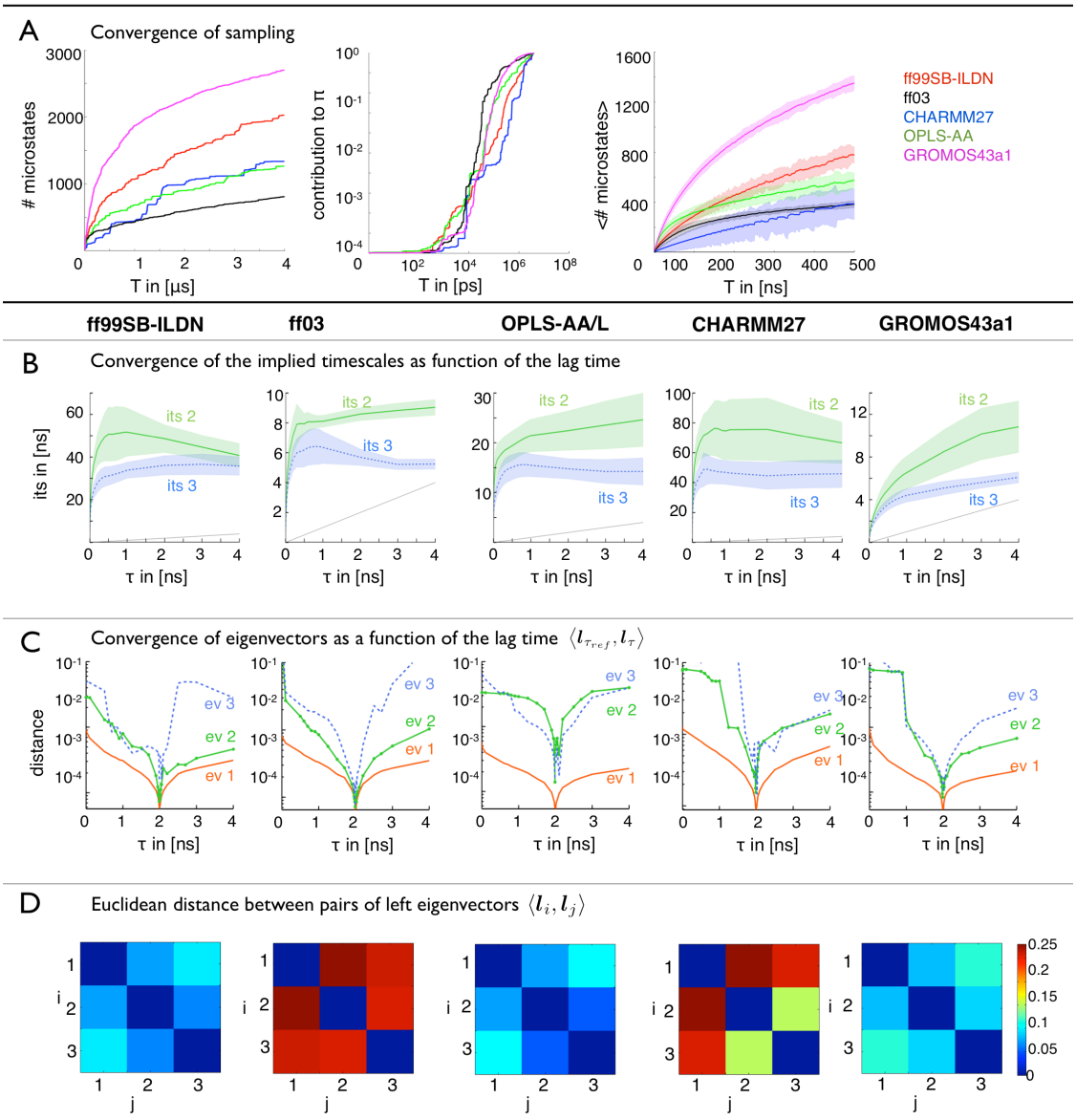

FIG. 5. Convergence checks for the simulation and Markov state models of the peptide deca-alanine. **A:** Convergence of the MD sampling with respect to the total simulation time T. **B:** Convergence of the implied timescales of the MSM. **C:** Convergence of the left eigenvectors of the MSM. **D:** Euclidean distance of pairs of the left eigenvectors of the MSM at lag time $\tau = 2$ ns.

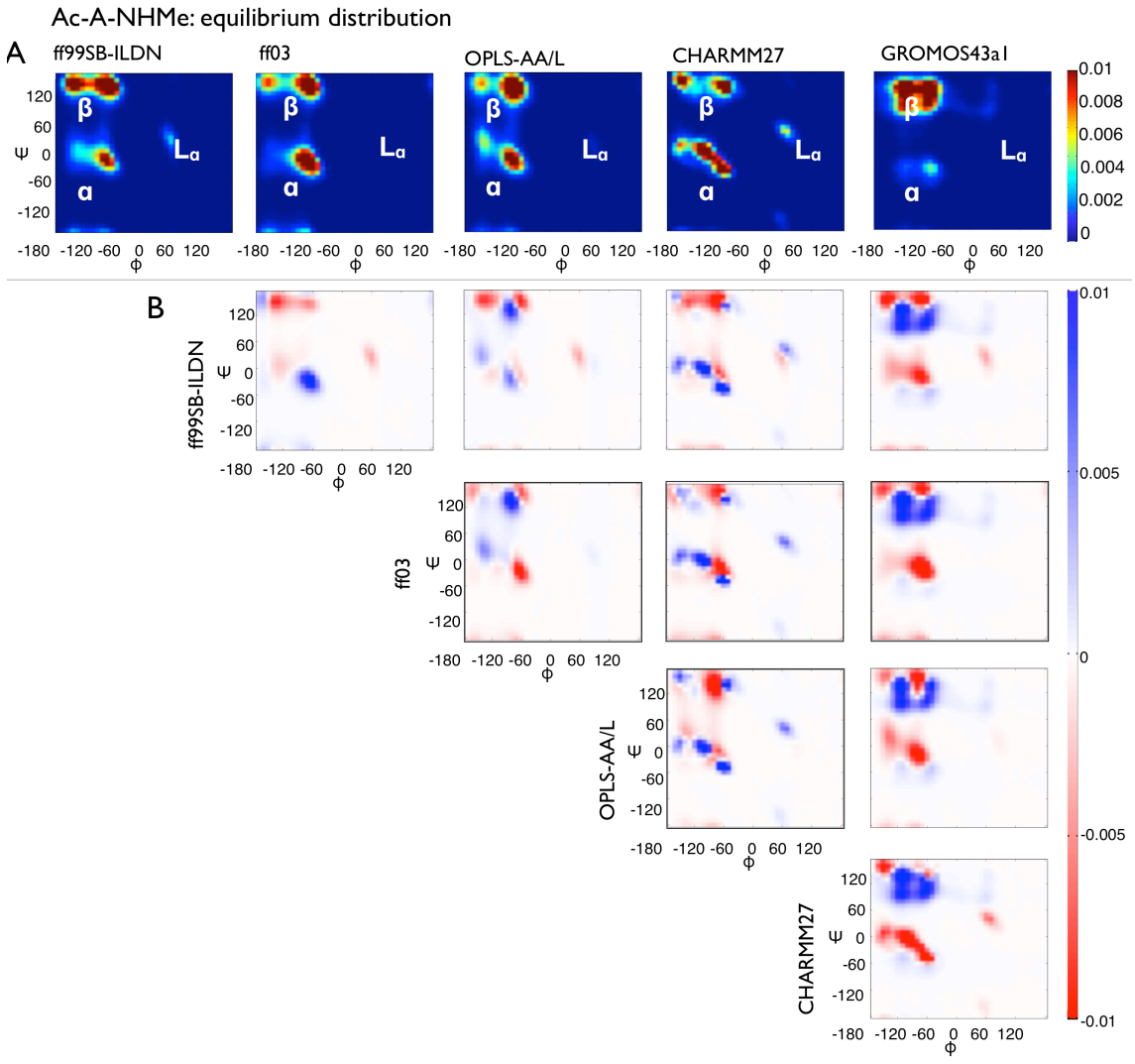

FIG. 6. Stationary process / equilibrium distribution of Ac-A-NHMe. **A:** Equilibrium distribution in the ϕ-ψ plane for the five force fields as estimated from the MSMs with lag time $\tau = 50ps$ (exception GROMOS43a1 $\tau = 10ps$). **B:** Pair-wise difference plots.

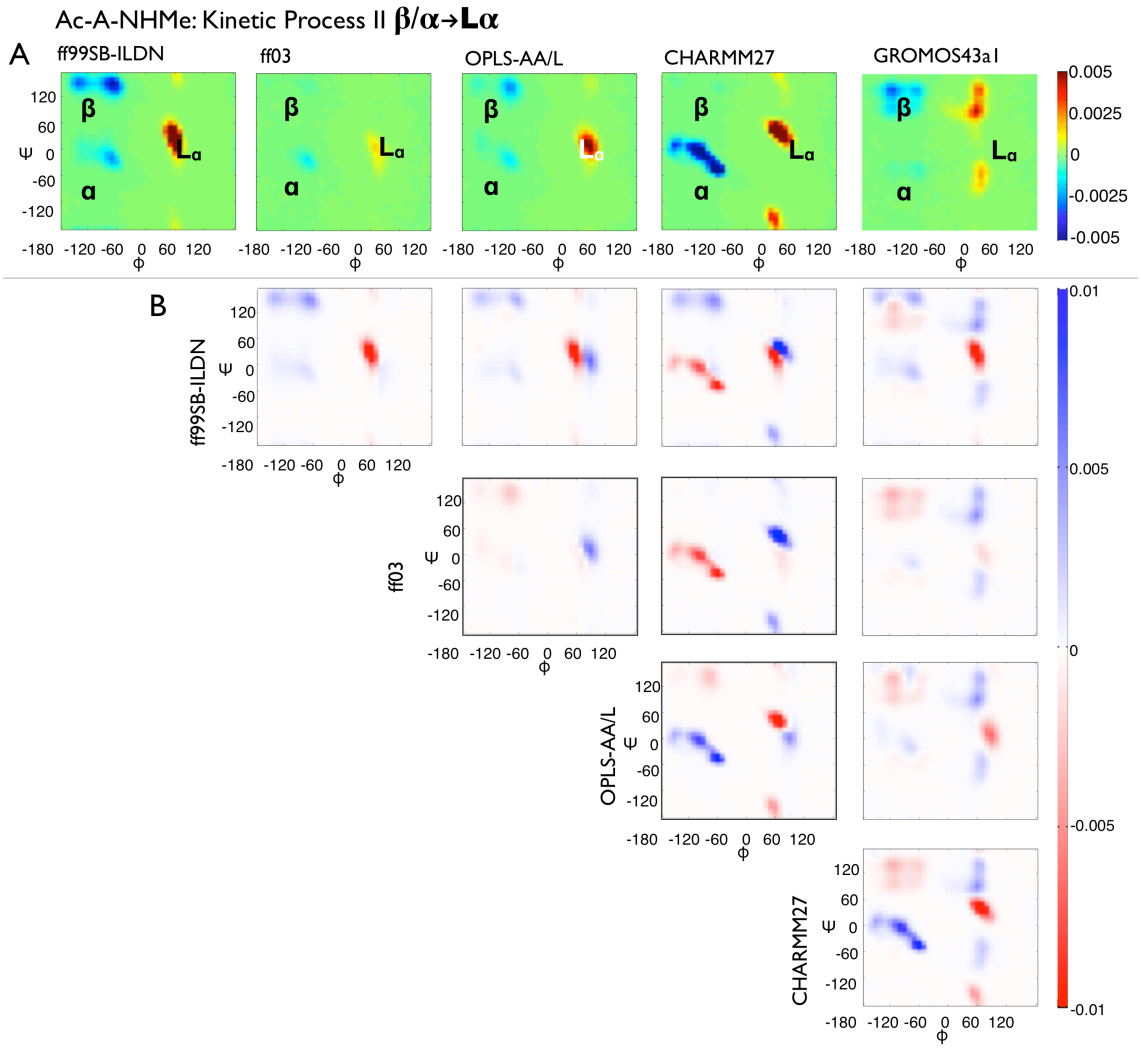

FIG. 7. Kinetic process which mediates the transition from the joint α and β conformations to the Lα conformation in Ac-A-NHMe. **A:** Kinetic process as estimated from the MSMs with lag time $\tau = 50ps$ (exception GROMOS43a1 $\tau = 10ps$). **B:** Pair-wise difference plots.

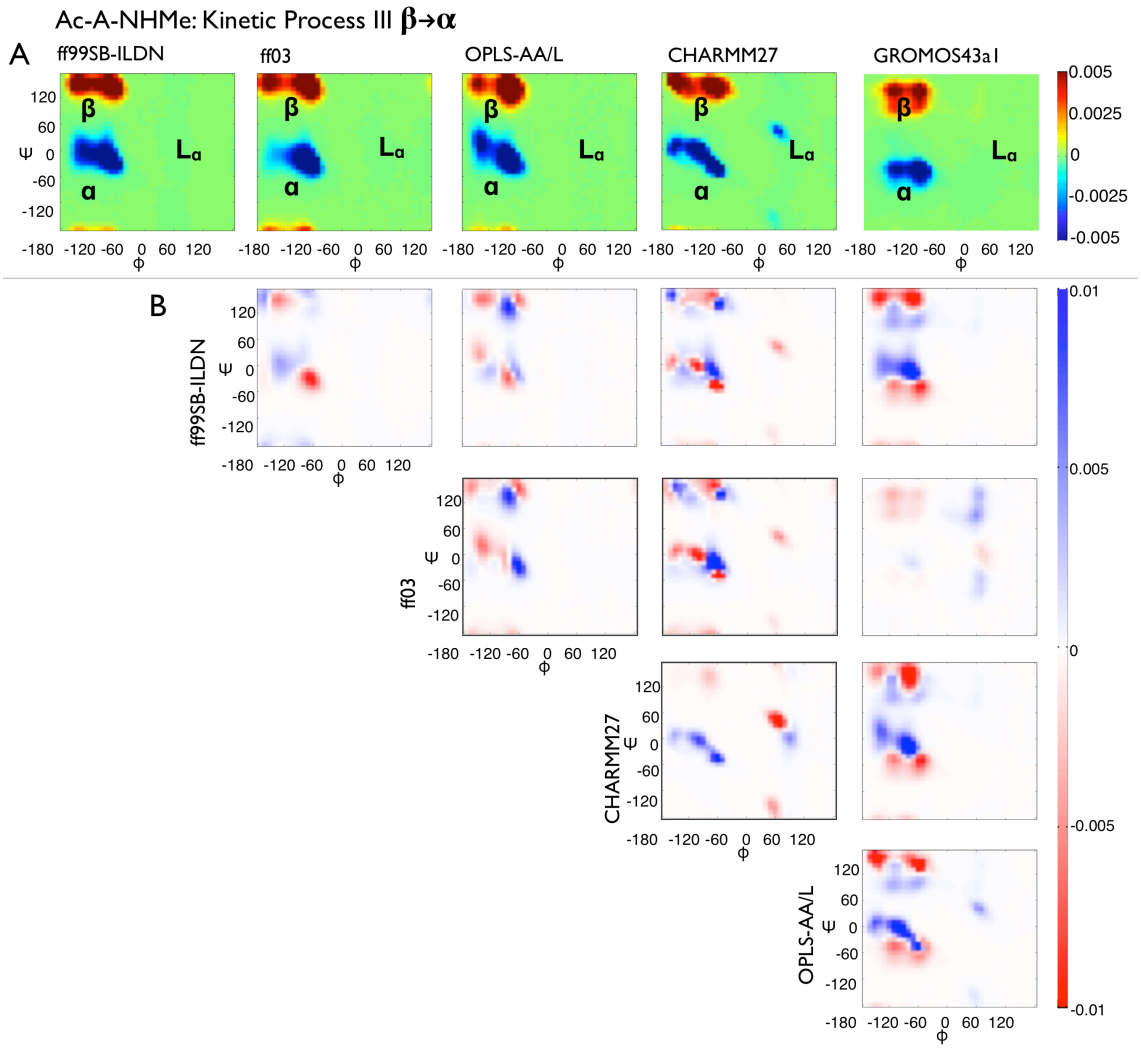

FIG. 8. Kinetic process which mediates the transition from the α conformation to the β conformation in Ac-A-NHMe. **A:** Kinetic process as estimated from the MSMs with lag time $\tau = 50ps$ (exception GROMOS43a1 $\tau = 10ps$). **B:** Pair-wise difference plots.

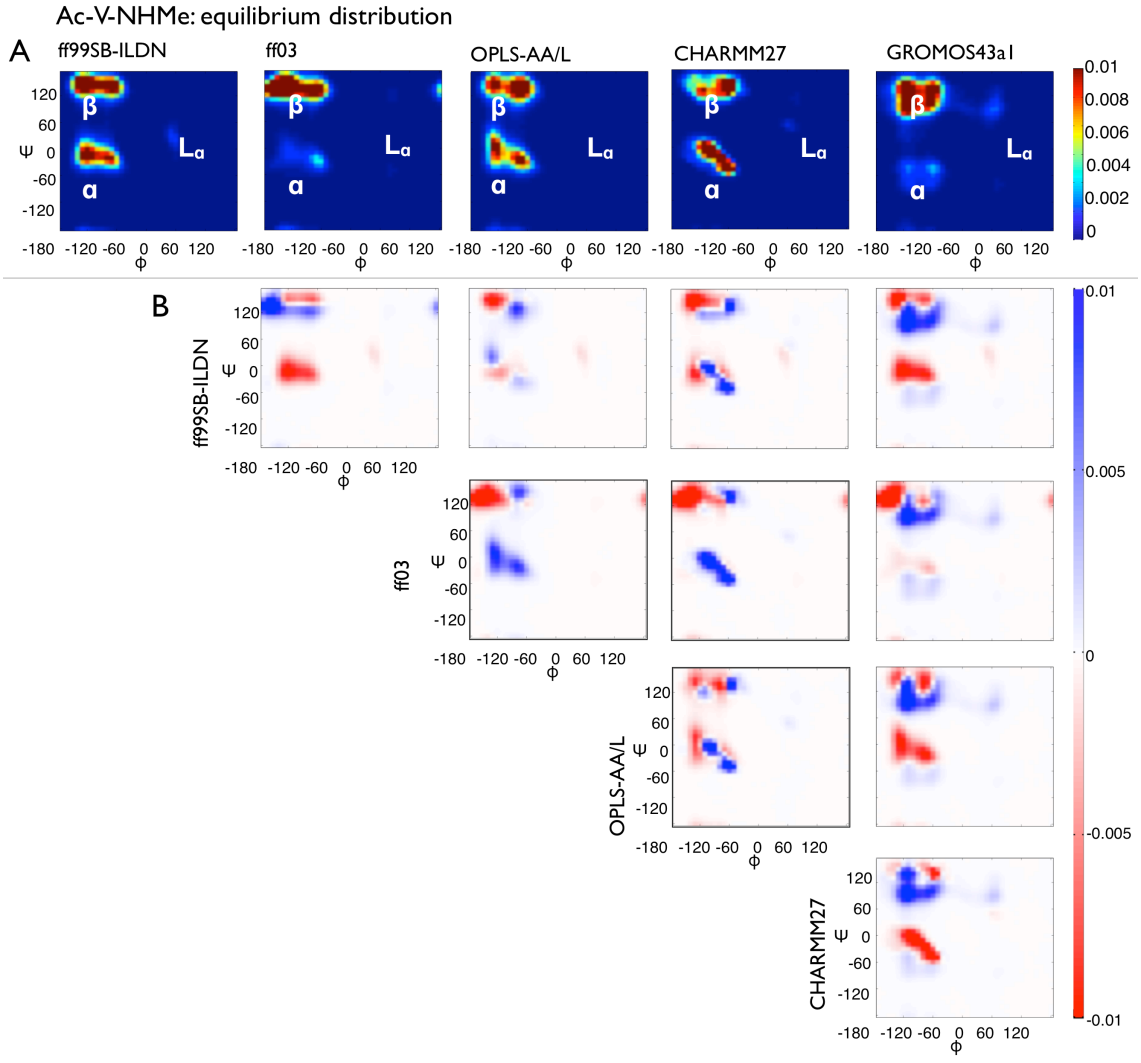

FIG. 9. Equilibrium distribution of Ac-V-NHMe. **A:** Equilibrium distribution in the ϕ-ψ plane for the five force fields as estimated from the MSMs with lag time $\tau = 50ps$. **B:** Pair-wise difference plots.

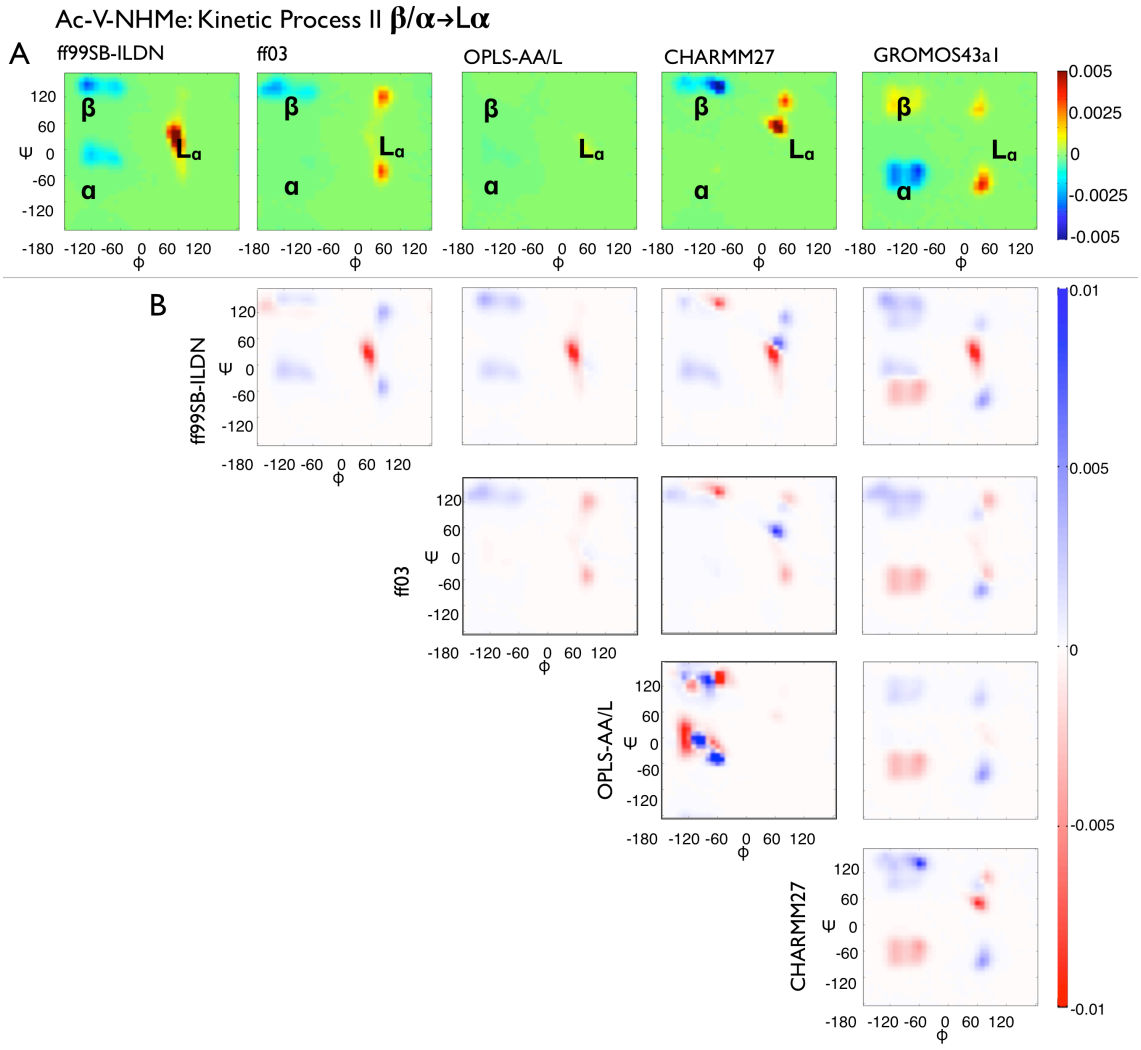

FIG. 10. Kinetic process which mediates the transition from the joint α and β conformations to the Lα conformation in Ac-V-NHMe. **A:** Kinetic process as estimated from the MSMs with lag time $\tau = 50ps$. **B:** Pair-wise difference plots.

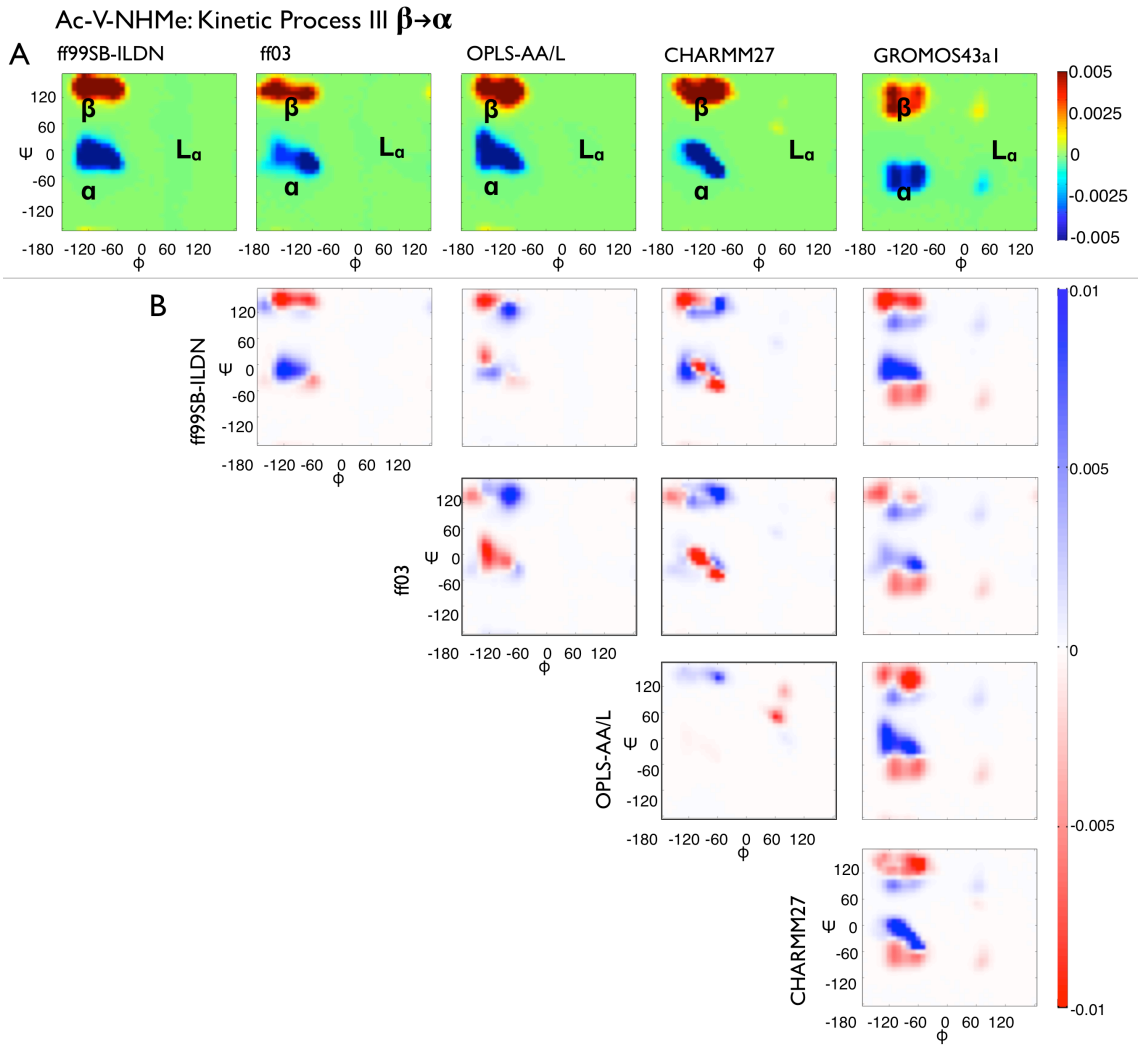

FIG. 11. Kinetic process which mediates the transition from the α conformation to the β conformation in Ac-V-NHMe. **A:** Kinetic process as estimated from the MSMs with lag time $\tau = 50ps$. **B:** Pair-wise difference plots.

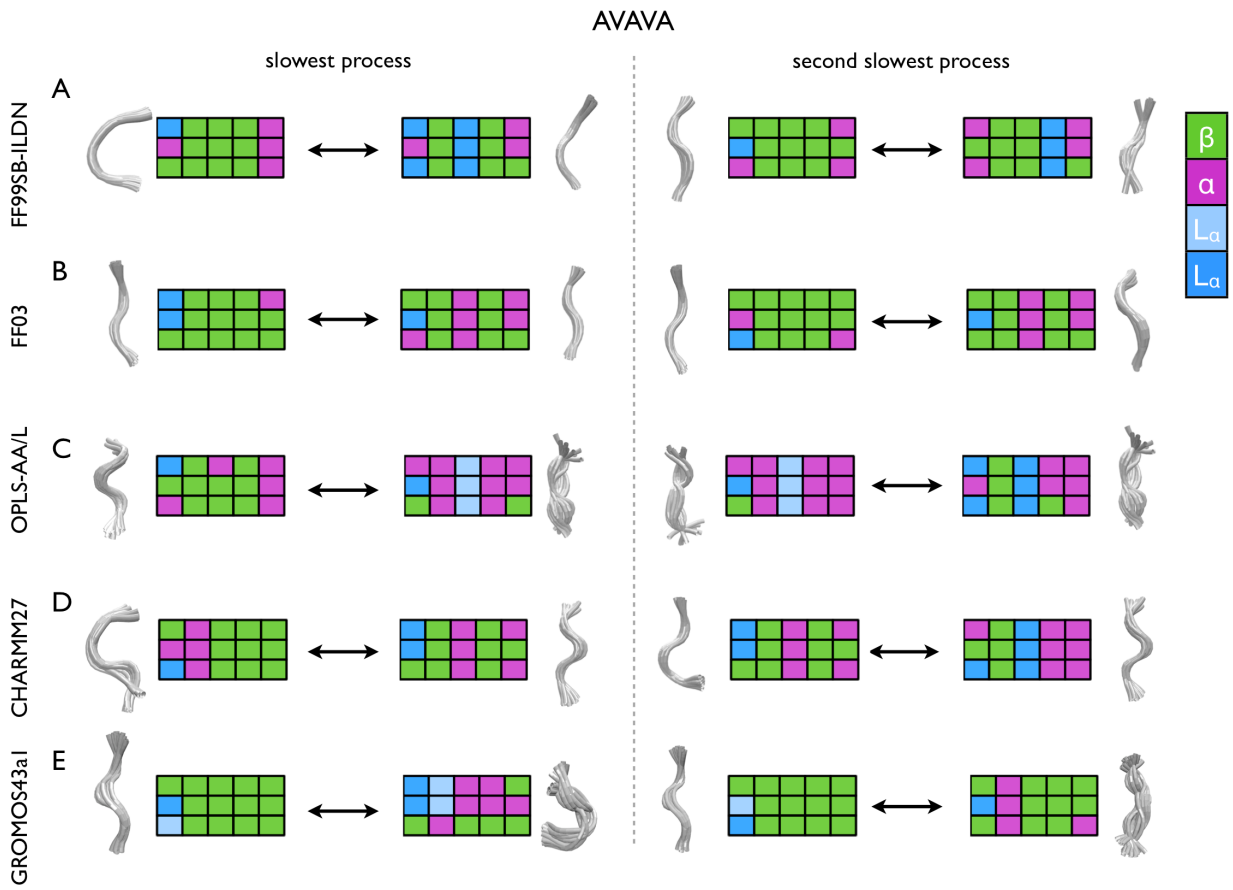

FIG. 12. Slowest kinetic processes in AVAVA visualized as transitions between backbone conformations. The plot shows the backbone conformations of representative microstates (which are kinetically connected by the process) as a sequence of colored boxes and as structural representative bundles.

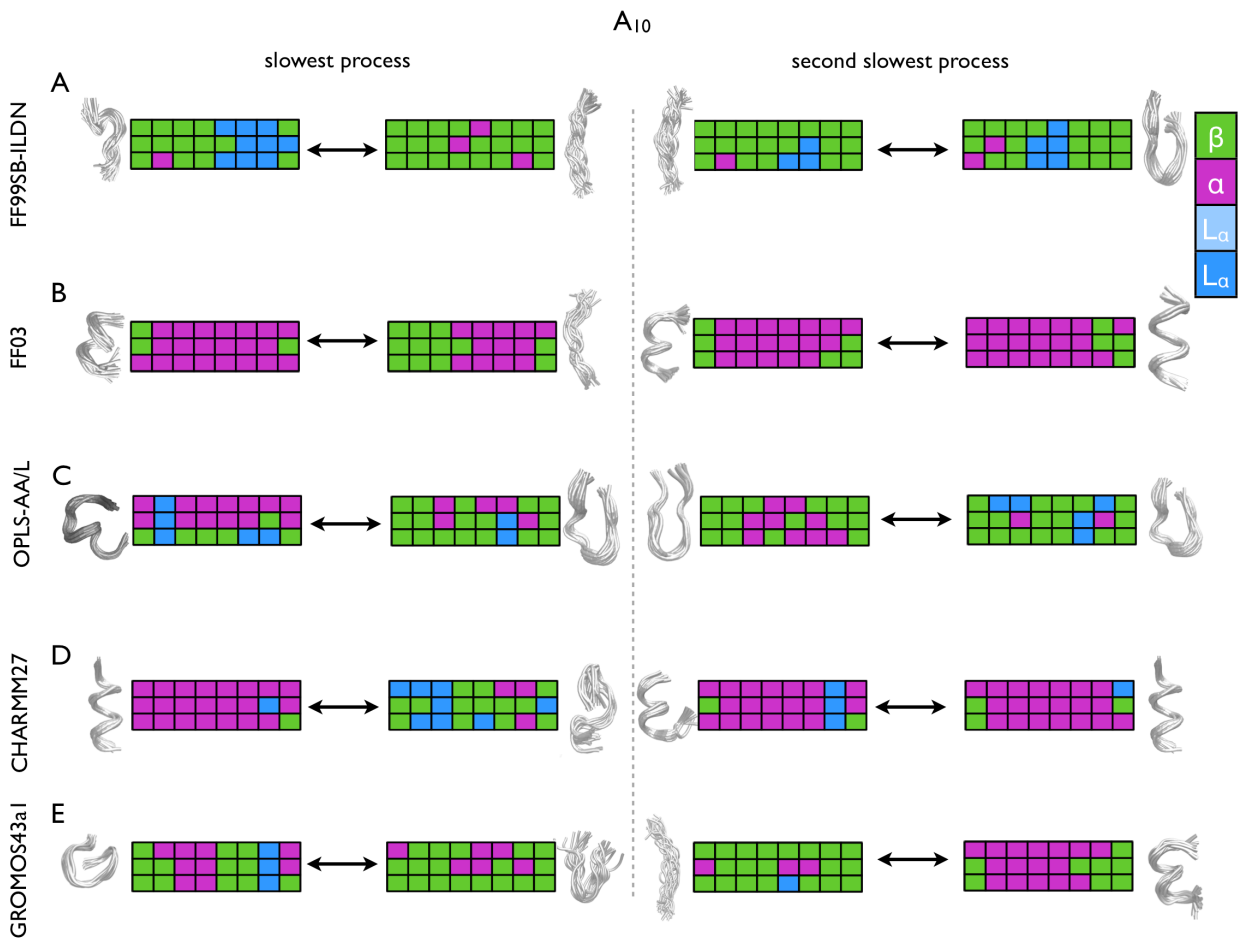

FIG. 13. Slowest kinetic processes in A_{10} visualized as transitions between backbone conformations. The plot shows the backbone conformations of representative microstates (which are kinetically connected by the process) as a sequence of colored boxes and as structural representative bundles.