Dynamic properties of force fields

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

(Received 24 September 2014; accepted 16 January 2015; published online 24 February 2015)

Molecular-dynamics simulations are increasingly used to study dynamic properties of biological systems. With this development, the ability of force fields to successfully predict relaxation timescales and the associated conformational exchange processes moves into focus. We assess to what extent the dynamic properties of model peptides (Ac-A-NHMe, Ac-V-NHMe, AVAVA, A_{10}) differ when simulated with different force fields (AMBER ff99SB-ILDN, AMBER ff03, OPLS-AA/L, CHARMM27, and GROMOS43a1). The dynamic properties are extracted using Markov state models. For single-residue models (Ac-A-NHMe, Ac-V-NHMe), the slow conformational exchange processes are similar in all force fields, but the associated relaxation timescales differ by up to an order of magnitude. For the peptide systems, not only the relaxation timescales, but also the conformational exchange processes differ considerably across force fields. This finding calls the significance of dynamic interpretations of molecular-dynamics simulations into question. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4909549]

I. INTRODUCTION

Currently, molecular dynamics (MD) simulations are arguably the only method by which the structure and the dynamics of large molecules can be studied simultaneously at atomistic resolution. Over the past decades, this method has hence become an invaluable tool for the investigation of biomolecular processes. However, the reliability of MD simulations has always been limited by two factors: (i) the length of the simulation which is accessible with the available computer resources and sampling algorithms (sampling problem), and (ii) the accuracy of the potential energy function used to calculate the forces in each MD simulation step (force field problem).

Advances in the sampling problem have often triggered improvements in force fields, as many force field deficiencies only became apparent with more extensive simulations. For example, when simulation times reached the microsecond regime,[1] and with the routine use of advanced sampling techniques, such as replica-exchange MD (REMD),[2,3] it was discovered that the secondary structure propensities of biomolecular force fields did not match the experimental data.[4,5] For different force fields, the α-helix conformation of overall secondary structures in polypeptides had relative probabilities ranging from 13% to 98%.[5] This suggested sizable differences in the exploration of the Ramachandran plane across different force fields and hence a need to reparametrize the ϕ- and ψ-backbone torsion angles. Moreover, the rotamer distributions of the χ_1 side-chain torsion angles varied considerably across force fields.[6,7] The conformational preferences of the side chain can in fact influence the equilibrium distribution of backbone torsion angles.[6,7] The differences in the structural distribution of individual torsion angles might explain why force fields differ in their ability to model the β-sheet formation[8] or the helix-coil transition.[9,10]

These findings have led to a series of reparametrizations in all of the major protein force field families. The ϕ- and ψ-torsion angles were reparametrized in the OPLS-AA/L[11] and in the AMBER force field,[7,12] whereas in CHARMM cmap, a grid-based energy correction for the ϕ-ψ plane, was introduced.[13,14] More recently, updated parameters for the χ_1 torsion angle were published for the AMBER and the CHARMM force field families.[10,15–17] By contrast, the GROMOS force field was reparametrized to reproduce the free enthalpies of hydration of model compounds,[18] leading to an update of the non-bonded interactions, in particular the partial charges. The rational behind this approach is that even when local interactions are modeled accurately, it is the relative free energy of solvation between polar and non-polar environments which drives processes such as protein folding (hydrophobic collaps) or membrane formation. Most of the reparametrized force fields exhibit a significantly improved agreement with experimental data which reports on structural properties of the molecule,[19,20] with some force fields even reaching experimental accuracy.[16]

We have now reached a point where the timescales accessible by direct MD simulation permit extensive sampling of small peptides. Thus, we can move beyond computing equilibrium populations of conformations by means of enhanced sampling techniques and compute the transition rates between these conformations. In the last few years, the dynamics of various molecular processes have been studied via extensive simulations. Examples are the study of protein and peptide folding using ultra-long trajectories[21,22] or swarms of short trajectories with Markov models.[23–26] Other methods which

a)F. Vitalini and A. S. J. S. Mey contributed equally to this work.
b)Electronic mail: frank.noe@fu-berlin.de
c)Electronic mail: bettina.keller@fu-berlin.de

were successfully used to explore slow processes and rare events are the milestoning approach[27] and transition path sampling.[28]

To understand why force fields, which give rise to the same (correct) folded state, differ in the underlying folding dynamics, one needs to consider the experimental data used for force field parametrization. These data are primarily of thermodynamic nature, i.e., they are ensemble averages of an observable O defined as

$$\langle O \rangle = \int_X \pi(x) O(x) \mathrm{d}x, \tag{1}$$

where $\langle \ldots \rangle$ denotes the expectation value, $\pi(x)$ the equilibrium probability density, and $O(x)$ is the value of the observable at a particular conformation x. The integral is evaluated over the entire conformational space X. For most observables, the integral in Eq. (1) is dominated by $\pi(x)$. The highly populated minima in the potential energy surface have a large impact on the observable averages, while $\langle O \rangle$ is less sensitive to the low population in the barrier regions, and hence to the relative height of the barriers. Also, entropic effects such as the shape and the extent of a minimum have a minor influence of the value of the integral. By contrast, the expectation value of dynamic properties (rates, timescale correlation functions) additionally depends on the transition density $\rho(x, y, \tau)$, which is extremely sensitive to the barrier heights and entropic effects. In other words, a parametrization against thermodynamic data ensures that the relative free energies of the minima are correct, but errors in the barrier heights or the shape of the minimum are likely to go unnoticed.

Besides barrier heights, three other aspects are thought to be decisive for an improved description of the folding dynamics: (i) the extent of the conformational space sampled by the unfolded protein,[29–31] (ii) the intrinsic stability of secondary structure elements,[32] and (iii) the cooperativity between the formation of the secondary structure elements.[20] However, a tool is missing which allows to link these rather global descriptors to the individual force field terms.

We suggest to use Markov state models (MSMs)[33–39] to characterize the dynamics induced by different force fields. These models approximate the conformational dynamics of a molecule as a jump process between (small) states in its conformational space. The resulting transition probabilities constitute the elements of the MSM transition matrix. To compare the dynamics induced by different force fields, we extract the dominant eigenvectors and associated eigenvalues of the MSM transition matrix. This has several advantages over a direct comparison of the matrix elements. First, the dominant eigenvectors contain information on the position of the minima and barriers of the potential energy surface, while the associated eigenvalues report on the timescale of the conformational exchange across these barriers. These two aspects cannot easily be disentangled when analyzing the transition probabilities directly. Second, given a valid MSM, the eigenvectors are independent of the lag time of the model, whereas the transition probabilities vary with the lag time. Third, the eigenvalues can be converted into relaxation timescales (or relaxation rates) which are directly linked to the timescales measured in correlation or pump-probe experiments.[40,41] Finally by analyzing the dominant eigenvectors, the comparison is restricted to the slow dynamical processes of the molecule, i.e., those processes which can be measured in experiment.

The paper is organized as follows. After a brief review of the theory of Markov state models, we compare the dynamics induced by various force fields on the level of a single residue using alanine and valine examples. We use capped amino acids (Ac-A-NHMe, Ac-V-NHMe) as minimal segments to model the dynamics of an amino acid residue in a protein or peptide chain.[33,42–46] Then, we assess whether the observed differences in the dynamics of single residues explain the differences in the dynamics of short peptides consisting of alanine and valine (AVAVA, A_{10}). A discussion of the implications of these results on the development of new force fields can be found at the end of this contribution.

We have chosen representatives of each of the four major force field families, focusing on those force field versions which have previously been used for the construction of Markov state models: ff99SB-ILDN,[15] OPLS-AA/L,[11] CHARMM27 with cmap,[13] and GROMOS43a1.[47,48] Additionally, we have included the AMBER force field ff03[49] because it shows a pronounced, albeit unphysical, helix-coil transition for the peptide deca-alanine. Note that ff99SB-ILDN differs from ff99SB only by a reparametrization of residues isoleucine (I), leucine (L), aspartic acid (D), and asparagin (N). The systems used in this studies do not contain any of these residues, and hence the results of ff99SB-ILDN should be equivalent to results which would have been obtained with ff99SB. Parameters which could influence the dynamics—other than the force fields—were identical in all of the simulations (integrator, integration time step, thermostats, constraints, etc.).

II. THEORY

In the following, we summarize the salient points of the theory of MSMs, which has been described in detail elsewhere.[33,36,37] In MSMs, the conformational space of a molecule is discretized into N states. The dynamics in this state space is determined by a transition matrix $T(\tau)$, whose elements represent conditional probabilities of finding the molecule in a conformation which belongs to state j, given that it was in state i in the previous time step

$$t_{ij}(\tau) = \mathbb{P}(x_{t+\tau} = j | x_t = i). \tag{2}$$

The time step τ of these transition probabilities is called the *lag time* and is typically chosen to be much larger than the time step of a MD simulation. Equation (2) defines a Markov process, i.e., the probability of finding the system in state j at time $t + \tau$ only depends on the current microstate ($x_t = i$), and not on the previous history of the process. The transition probabilities $t_{ij}(\tau)$ can be estimated from MD trajectories by counting the observed transitions from state i to state j in a trajectory.[37]

For systems with ergodic and reversible dynamics, the transition matrix is decomposable into a complete set of real-valued (left) eigenvectors l_i and associated eigenvalues $\lambda_i(\tau)$ which are defined by

$$l_i^\top T(\tau) = \lambda_i(\tau) l_i^\top, \tag{3}$$

where $l_i^\top$ denotes the transpose of l_i. For all transition matrices, the eigenvalue with the highest absolute value is $\lambda_1 = 1$. All other eigenvalues are guaranteed to be smaller in absolute value: $|\lambda_i(\tau)| < 1$ for $i > 1$. This eigendecomposition is extremely useful because the salient characteristics of the conformational dynamics can be understood by interpreting the dominant (high-lying) eigenvectors of the transition matrix and the associated eigenvalues.[37]

To illustrate this, let us consider an ensemble of systems as it relaxes from some out-of-equilibrium distribution $p(0)$ towards the equilibrium distribution π. The evolution of the ensemble probability distribution can be written as superposition of the eigenvectors l_i of the transition matrix governing the dynamics of the individual systems in the ensemble

$$p(t) = \sum_{i=1}^{N} c_i \lambda_i^{t/\tau}(\tau) l_i. \tag{4}$$

The expansions coefficients c_i are determined by the initial distribution $p(0)$. The eigenvectors can be interpreted as dynamic modes which are associated to time-dependent amplitudes

$$a_i(t) = c_i \lambda_i^{t/\tau}(\tau) = c_i \exp\left(-\frac{t}{t_i}\right). \tag{5}$$

This implies that the amplitude of the first eigenvector l_1 is constant, and that the amplitudes of all other eigenvectors decay exponentially,

$$\lambda_i^n(\tau) = \begin{cases} 1 & i = 1 \\ \exp(-\kappa_i n \tau) & i > 1. \end{cases} \tag{6}$$

While the eigenvalue $\lambda_i(\tau)$ depends on the chosen lag time τ, the rate of decay κ_i does not. It is given as

$$\kappa_i = -\frac{\ln \lambda_i(\tau)}{\tau} = \frac{1}{t_i} \quad \forall i > 1. \tag{7}$$

Frequently, instead of the relaxation rate, the relaxation time or implied timescale t_i is reported, which is simply the inverse of κ_i.

Equations (4) and (5) imply that, after an initial relaxation time, the dynamics of the ensemble probability distribution is dominated by dynamic modes (eigenvectors) which are associated to eigenvalues with absolute values close to one. These so-called dominant eigenvectors have small decay rates and large implied timescales. They represent the slow processes of the conformational dynamics of the molecule. Due to the fast decay of the non-dominant modes, a reduced representation of the conformational dynamics of a molecule in terms of its dominant eigenvectors often yields an accurate and sufficiently detailed picture. This is particularly true, if there is a gap between the dominant eigenvalues and the non-dominant eigenvalues.

The first eigenvector l_i is called the stationary process and this eigenvector is equal to the equilibrium probability distribution: $l_1 = \pi$. As a consequence, l_1 does not have negative entries. It is the only mode which does not decay in the limit $t \to \infty$. This reflects the physical expectation that any initial probability distribution will relax towards the equilibrium distribution if the ensemble is not perturbed during the relaxation process. All other eigenvectors have positive and negative entries, and entries close or equal to zero. Effectively, a given eigenvector l_i assigns a value to each state in the conformational space. The vector can be interpreted as a dynamic process which mediates the conformational exchange between set of states with negative values and sets of states with positive values. States to which the eigenvector assigns a value of zero are not affected by this particular process (Fig. 2). The kinetic exchange between the two sets of states does not necessarily occur via a single, specific transition state. Rather, all possible paths which connect one set with the other contribute to the conformational exchange represented by l_i. Because of this, the relaxation rates cannot be interpreted as transition rates across a potential energy barrier. An analysis of the dominant eigenvectors yields information on long-lived conformational states and on free-energy barriers between them (but not necessarily on the transition states). Additionally, by comparing the eigenvectors of the MSMs of two molecules, one can assess the differences in the dynamics of these molecules.

Equation (7) holds for any Markov process. However, representing the conformational dynamics of a molecule as a jump-process between conformational states, never yields a dynamics which is completely Markovian. As a consequence, the relaxation rates $\hat{\kappa}_i(\tau)$ and relaxation times $\hat{t}_i(\tau)$, which are estimated from MD data, vary with the lag time. The deviation from Markovian dynamics decreases with increasing lag times which is reflected in the lag time dependence of the relaxation times. In practice, one estimates the implied timescales $\hat{t}_{i>1}$ of the slowest processes for different lag times and chooses a value of τ for which the τ-dependence of $\hat{t}_i$ has become negligible (Figs. 2(b), 3(b), 4(b), and 5(b) in the supplementary material[50]). For completeness, we note that the approximation quality of a MSM also sensitively depends on the chosen discretization of the conformational space.[51] However, for the small systems investigated in this paper, discretization is not a bottleneck.

III. METHODS

A. Simulations setup

We performed all-atom MD simulations in explicit water of acetyl-alanine-methylamide (Ac-A-NHMe), acetyl-valine-methylamide (Ac-V-NHMe), a penta-peptide with sequence AVAVA, and a deca-peptide composed of ten Alanine residues (A_{10}). All simulations were carried out with GROMACS 4.5.5 simulation package,[52] in an NVT ensemble at 300 K. The atom positions of the solute were saved to file every 1 ps. Each system was simulated in five different force fields: AMBER ff99SB-ILDN, AMBER ff03,[49] OPLS-AA/L,[11] CHARMM27,[13] and GROMOS43a1.[47,48] The water model was chosen according to the one used for the force fields validation, i.e., TIP3P[53] for AMBER ff99SB-ILDN,[12,15] AMBER ff03,[49] OPLS-AA/L[11,54] and CHARMM27,[13] and SPC[55] for GROMOS43a1.[47,48] For each of these setups, the aggregated simulation time was 4 μs, adding to a total simulation time of 80 μs. The convergence of the sampling of the

conformational space has been tested (data shown in the supplementary material,[50] Figs. 2(a), 3(a), 4(a), and 5(a)). Further simulation details can be found in Sec. I of the supplementary material.[50]

B. Markov state model construction

The Markov state models were constructed on the conformational space spanned by the ϕ- and ψ-backbone torsion angles, which have been proven to be useful reaction coordinates to capture short-peptides dynamics.[35] For the capped amino acids, the ϕ-ψ-plane was discretized by a regular grid of $36 \times 36 = 1296$ microstates, where each state had a size of $10° \times 10°$. For larger peptides, such a high resolution is computationally not tractable. Therefore, the ϕ-ψ-planes of residues 2–5 of AVAVA have been discretized into four bins (Fig. 1 in the supplementary material[50]), where each bin corresponds to a peak of the equilibrium distribution in the ϕ-ψ-plane. The N-terminal residue exhibits a different sampling of the ϕ-ψ-plane, leading to a grid with six bins for this residue (left graphs in Figs. 1(a) and 1(b)). Each possible configuration of bins along the peptide chain represents a microstate, resulting in $6 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 1536$ microstates for the MSMs of this peptide. We verified that this discretization is able to separate the local maxima of the ϕ-ψ-equilibrium distribution in the used force fields (Fig. 1 in the supplementary material[50]). The MSMs of deca-alanine were constructed on the ϕ- and ψ-backbone torsion angles of residues 2–8. The ϕ-ψ-plane was discretized into three bins (right graphs in Figs. 1(a) and 1(b)), where the Lα bins correspond to the combined bins two and three of the four-state discretization used for AVAVA. This discretization yields $3^8 = 6561$ microstates.

The MSMs were constructed by counting the transitions between microstates. For detailed description of the estimation procedure, see supplementary material.[50] The eigenvalues and eigenvectors of the MSM transition matrices were calculated and used for comparing dynamic properties of the force fields. Convergence checks for the MSMs, including implied timescale plots, are shown in Figs. 2–5 in the supplementary material.[50]

C. Autocorrelation function

Fits to autocorrelation functions of interesting molecular observables are often used in order to determine relaxation rates or timescales. It can be shown that all relaxation timescales of the molecule, which can be approximated using a MSM, will appear in the autocorrelation function, although some may have negligible amplitudes and can thus not be retrieved by a multiexponential fit.[40,41] In order to demonstrate this problem, we compute autocorrelation functions (ACF) of C-α RMSD with respect to an idealized α-helix for both peptides and estimate relaxation timescales from multiexponential fits. In the literature, an idealized α-helix is defined as backbone dihedral angles of $(\phi,\psi) = (-62°, -41°)$. In the case of A_{10}, we obtained a reference structure that is as close to the ideal helix as possible, by selecting a reference where all the eight non-terminal residues' dihedral pairs were in a 20° radius from the ideal structure. For AVAVA, using the above helix definition, no reference structure in any trajectory can be found; therefore, the peptide never forms an ideal helix. Instead a reference was selected from the α-helical state of the MSM discretization for the three non-terminal residues. With respect to each reference structure, RMSD trajectories were computed, using the `g_rms` tool of GROMACS.[52] From all RMSD trajectories, we calculate the ACF for each forcefield, respectively. For each forcefield and peptide, we fit a double exponential to the ACF, except in the case of CHARMM27 in the analysis of the A_{10} data, where a single exponential fit describes the behavior of the ACF better. Figure 6 shows the logarithm of the ACF and the fit, with the time constants obtained from the fits listed in Table II.

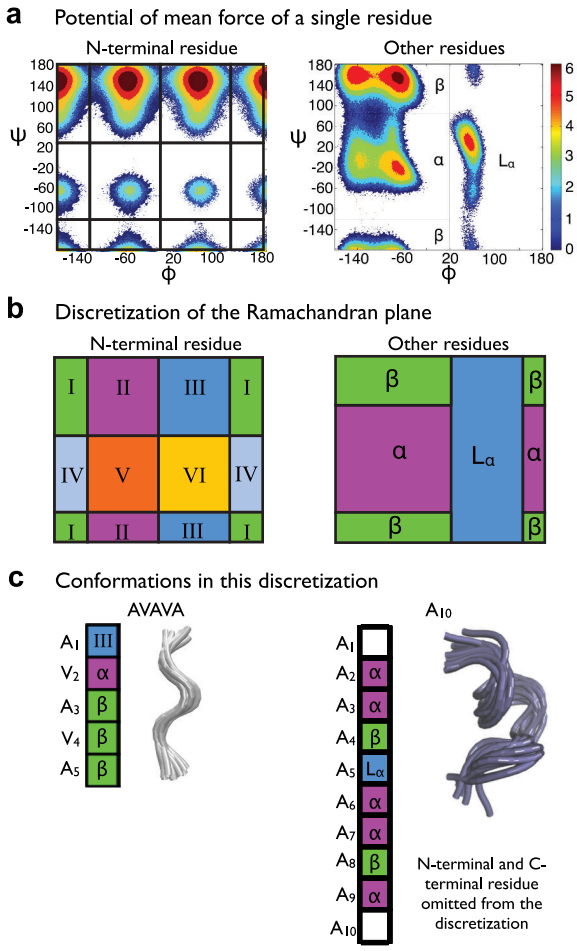

FIG. 1. Discretization for an amino acid in a peptide. Panel (a) shows the logarithm counts of the backbone torsion angles distribution of an amino acid at the N-terminus (left) or within the peptide sequence (right). In (b) the discretization for both distributions is given. The states are defined to capture the main secondary structure features of the Ramachandran-plane. A different color is associated to each state: α is color-coded purple, β is green, and Lα is blue. For the N-terminus amino acid (right), a color-scheme is given, however, the secondary structure interpretation is lost. Panel (c) shows a discrete state for a peptide (AVAVA left and A_{10} right) as colored string, where each square represent the state of a residue in the sequence. For each state, an example of conformations is provided as a bundle.

IV. RESULTS AND DISCUSSION

A. Capped amino acids

Ac-A-NHMe, often referred to as alanine dipeptide, is an alanine molecule whose N-terminus is capped by an acetyl

group, and its C-terminus is capped by a methyl amide group. Its dynamics can be described in terms of the ϕ- and ψ backbone torsion angles. The equilibrium distribution in the $\{\phi - \psi\}$-space (Ramachandran plot, Fig. 1(a)) can be derived by considering the steric interactions of the carbonyl groups and the side chain.[56] Characteristic for all amino acids, with the exception of capped Proline and Glycine, are five population maxima (Fig. 1(a)): four in the $\phi < 0$ region ($\beta = \beta$, ppII, and $\alpha = \alpha_R, \alpha_P$) and one at positive values of ϕ (Lα).

We constructed MSMs for capped alanine and valine on a regular grid in the ϕ-ψ-space and extracted the equilibrium distribution as the first left eigenvector of the transition matrix of each MSM (first column of Figs. 2(a) and 2(b)). Qualitatively, all force fields capture the characteristic features of the Ramachandran plot. However, Figs. 2(a) and 2(b) also highlight the known differences[10,12,15,19,20,57,58] in the positions, relative height and shape of the population maxima between different force fields.

In particular, the relative population of the Lα-region of the Ramachandran plot varies considerably across the five force fields.

A low population in this region might be linked to an overpopulation of the α-helical conformation in larger peptides.[4,5] We find differences in the shape of the equilibrium distribution between ff03, ff99SB-ILDN, OPLS-AA/L on the one hand and CHARMM27 and GROMOS43a1 on the other. While in the first group the α- and β-maxima are modelled as doubly peaked maxima, CHARMM27 presents three maxima in the α-region, and GROMOS43a1 has four maxima in the β-region. Likewise, the barrier at around $\phi = 0°$ is pronounced in ff03, ff99SB-ILDN, and OPLS-AA/L, whereas there are low-lying transition states in CHARMM27 and GROMOS43a1. Difference plots of the equilibrium distributions are shown in the supplementary material[50] in Fig. 6 for Ac-A-NHMe and in Fig. 9 for Ac-V-NHMe.

Despite these differences in the minima and barriers in the ϕ-ψ space, the slow dynamic processes in this process are similar in all five force fields and both residues. We extracted the second and third left eigenvectors of each of the MSMs (Eq. (4)) and grouped them into two types of dynamic processes (second and third column in Figs. 2(a) and 2(b)). Overall, process II switches between regions with $\phi < 0°$, and regions with $\phi > 0°$. This process mediates the conformational exchange between the Lα-region and the combined α- and β-region. Process III on the other hand switches between the α- and the β-region by assigning values of opposite signs to

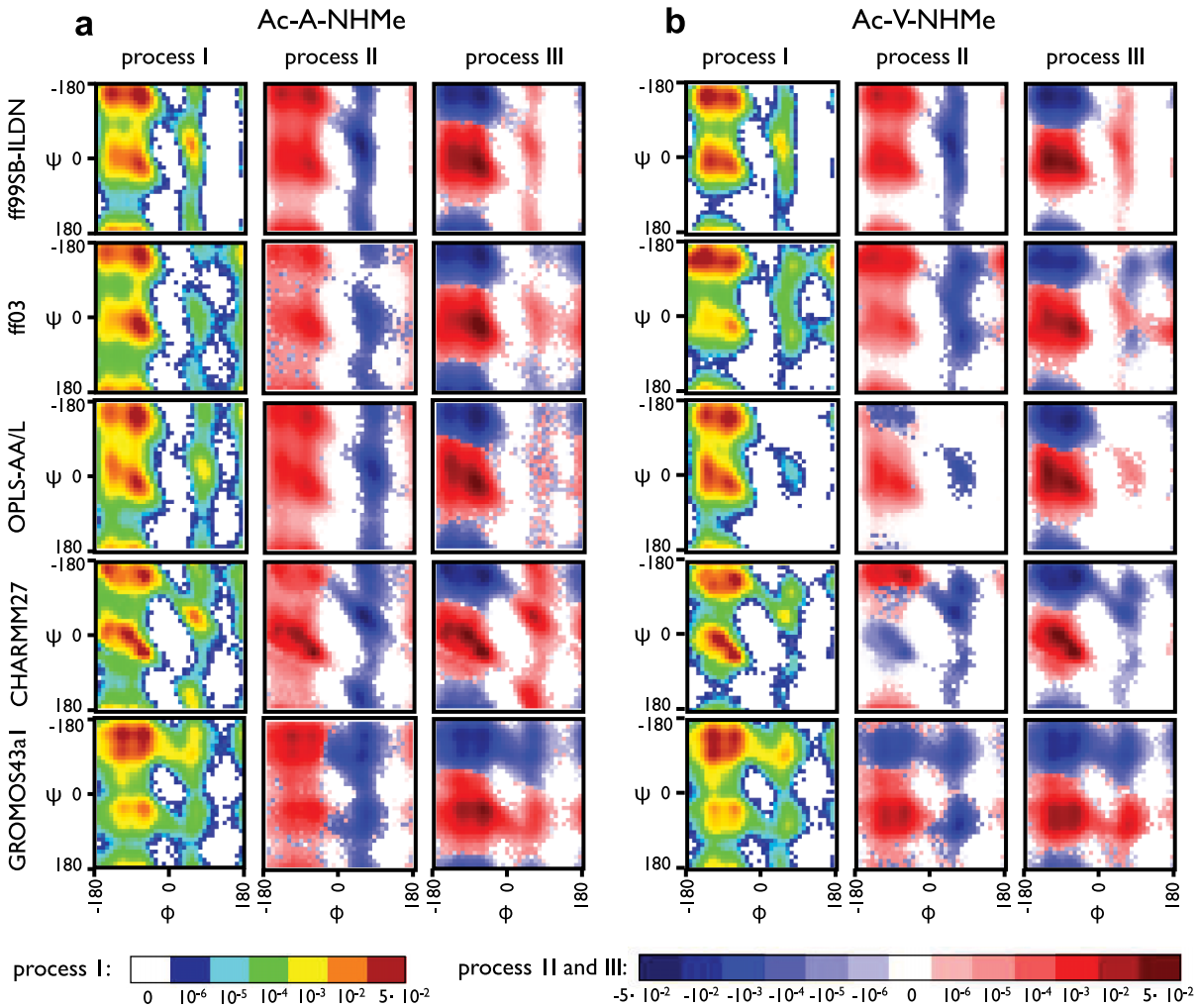

FIG. 2. Slowest process eigenvectors of alanine (a) and valine (b) dipeptides for all force fields sorted according to processes. In column one (process I), the equilibrium distribution for each force field is shown. Process II indicates a dynamic process along the ϕ coordinate; probability distribution is here transported across the energy barrier at $\phi = 0$, between the $\beta - \alpha$-regions and the Lα minimum. Process III indicates transitions along the ψ coordinate ($\beta \leftrightarrow \alpha$).

these areas. This process thus mediates the conformational equilibrium between the α- and β-regions.

While processes II and III are distinct in most of the Ac-A-NHMe simulations, we find some mixing of the processes in Ac-V-NHMe. For example, the eigenvectors classified as process II assign negative values to some parts of the α- or β-regions in OPLS-AA/L, CHARMM27, and GROMOS43a1 (Fig. 2). However, since the processes are dominated by the switching between regions with positive ϕ values and regions with negative ϕ values, the classification of these eigenvectors as process II is justified.

Likewise, all eigenvectors which predominantly mediate the dynamic exchange along the ψ-coordinate are classified as process III. (For difference plots of the two processes, see the supplementary material[50] Figs. 7, 8, 10, and 11 for Ac-A-NHMe and Ac-V-NHMe, respectively.) Panels (c) and (d) of Figs. 2 and 3 in the supplementary material[50] also show convergence tests for the eigenvectors.

The fact, that the eigenvectors can be classified as either process II or III, implies that, on the level of a single residue, the dynamics in all five force fields is dominated by a similar topology of the potential energy surface in the ϕ-ψ space (position of the barriers and positions and relative depths of the minima).

However, the corresponding relaxation rates differ by up to an order of magnitude across different force fields. Fig. 3 displays the relaxations rates associated to process II (blue line) and III (gray line) as speed-o-meters plots. Most notably, the order of processes II and III is not preserved across the force fields. In most simulations, process II is the slowest process, while process III is associated to a higher relaxation rate. However, for Ac-A-NHMe in ff03, Ac-V-NHMe in CHARMM27, and both capped amino acids in GROMOS43a1, process III is slower than process II (speed-o-meters with a red circle in Fig. 3). The dashed lines in Fig. 3 represent the statistical uncertainty of the relaxation rates, estimated as the standard deviation in a bootstrap sample analysis. (See supplementary material[50] for details on the bootstrap analysis.)

The rates can be interpreted as an order-parameter for the speed at which the force field samples the conformational space: the higher the rate, the faster the sampling. Fig. 3 reveals a large difference of the sampling speed between the force fields. For example, the rate of process II in Ac-A-NHMe is estimated to be less than 1 ns^{-1} in ff99SB-ILDN, while GROMOS43a1 assigns a rate of 85 ns^{-1} to the same process—a difference of two orders of magnitude. Moreover, the relative magnitude of the rates differs across force fields. For some simulation setups, the rates of process II and III are in the same range (e.g., Ac-A-NHMe with ff03 and CHARMM27, Ac-V-NHMe with ff99SB-ILDN and OPLS-AA/L), while in other force fields they differ considerably (e.g., Ac-A-NHMe with ff99SB-ILDN and OPLS-AA/L).

Based on the simulations of capped amino acids, it is not possible to identify a particular "fast" or "slow" force field. GROMOS43a1 yields exceptionally high rates for Ac-A-NHMe, but the rates for Ac-V-NHMe are comparable to those of other force fields. The relaxation rates and relaxation timescales (including standard deviations) for all simulations of Ac-A-NHMe and Ac-V-NHMe are summarized in Table I.

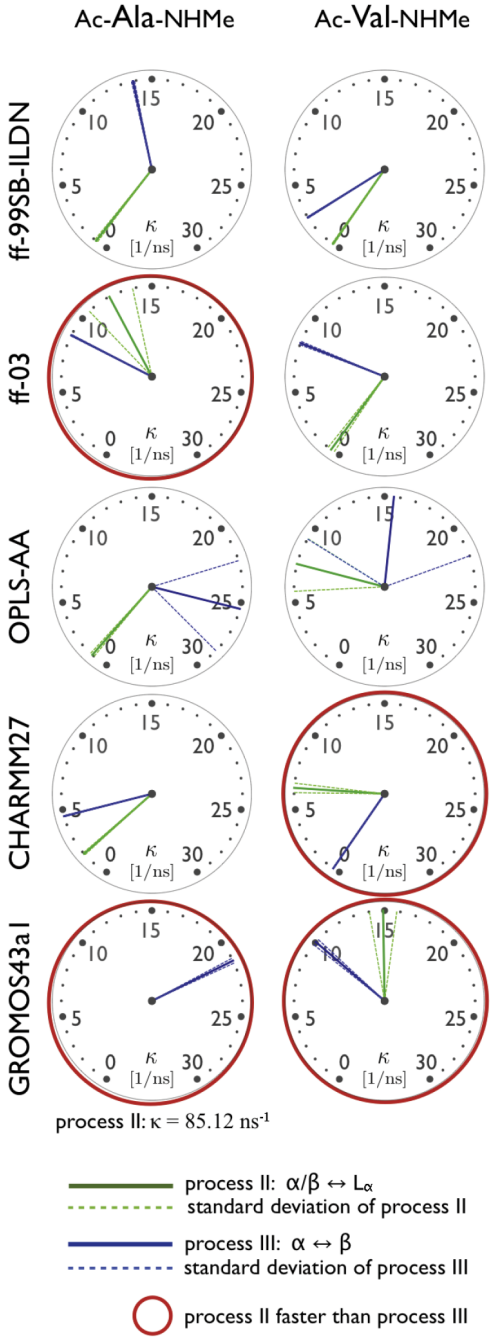

FIG. 3. Implied timescales rates for the slowest processes of alanine and valine dipeptides. The rates associated to the dynamic processes are plotted in a polar form to emphasize the different speed of the force fields. The transition $\beta/\alpha \leftrightarrow \alpha$, i.e., process II is shown in green; the transition $\beta \leftrightarrow \alpha$ in blue. Speed-o-meters surrounded by a red circle identify force fields for which the order of the processes is switched. Statistical uncertainty, estimated as the standard deviation in a bootstrap sample analysis, is identified with a dashed line; in some cases, it is order of magnitudes smaller than the associated timescale, thus difficult to visualize; in Ac-V-NHMe with OPLS-AA, the error lines of the two timescales overlap. Table I provides timescales and rates of bootstrap mean and standard deviation values for both capped amino acids in each force fields.

B. Peptides

To test how the difference in backbone torsion dynamics affects the overall dynamics of peptide systems, we performed a MSM analysis of short peptides, whose sequence included

TABLE I. Relaxation rates k_i and the relaxation timescales t_i of the second ($i = 2$) and third ($i = 3$) dynamic process of the simulated capped amino acids. Reported are mean and standard deviation from the bootstrap analysis.

System: Ac-Ala-NHMe				
	Process II		Process III	
Force field	κ_2 (ns^{-1})	t (ns)	κ_3 (ns^{-1})	t (ns)
ff99SB-ILDN	0.788 ± 0.104	1.269 ± 0.166	13.791 ± 0.095	$0.073 \pm 5\times10^{-4}$
ff03	12.008 ± 1.567	0.083 ± 0.011	8.720 ± 0.050	$0.115 \pm 6\times10^{-4}$
OPLS-AA/L	1.093 ± 0.146	0.915 ± 0.122	24.863 ± 3.094	0.040 ± 0.006
CHARMM27	1.841 ± 0.080	0.543 ± 0.024	4.568 ± 0.034	0.219 ± 0.002
GROMOS43a1	83.850 ± 1.093	$0.012 \pm 2\times10^{-4}$	21.350 ± 0.164	$0.047 \pm 3\times10^{-4}$
System: Ac-Val-NHMe				
	Process II		Process III	
Force field	κ_2 (ns^{-1})	t (ns)	κ_3 (ns^{-1})	t (ns)
ff99SB-ILDN	0.468 ± 0.061	2.139 ± 0.287	2.794 ± 0.047	0.358 ± 0.006
ff03	0.543 ± 0.236	1.840 ± 0.609	8.181 ± 0.111	0.122 ± 0.002
OPLS-AA/L	6.770 ± 1.750	0.148 ± 0.053	12.945 ± 6.404	0.077 ± 0.031
CHARMM27	6.370 ± 0.301	0.157 ± 0.007	0.449 ± 0.010	2.229 ± 0.051
GROMOS43a1	14.883 ± 0.873	0.067 ± 0.004	10.065 ± 0.219	0.099 ± 0.002

alanine and valine residues: a penta-peptide with sequence AVAVA and deca-alanine. Poly-alanine peptides have been used before for force field studies.[5,10]

The MSMs were constructed on the conformational space spanned by the ϕ- and ψ-backbone torsion angles, where the ϕ-ψ-plane of a residue was discretized into three (A_{10}) or four (AVAVA) states (Fig. 1(b)). These states can be mapped to the three regions in the Ramachandran plane: α, β, and Lα. Each peptide conformation can then be represented by a string in which the ith element represents the conformation of the ith residue. Fig. 1(c) explains how these strings are converted into the color-coded representations shown in Figs. 4 and 5. The terminal residues exhibit different dynamics than residues in the middle of a peptide chain and are often largely decoupled from the dynamics of the overall peptide chain. Therefore, we used a six-state discretization for the N-terminal residue of AVAVA (Fig. 1(b)) and excluded residue 1 and 10 of A_{10} from the MSM.

In Fig. 4, the slowest processes of AVAVA for the five force fields are shown as dynamic exchange processes between sets of conformations. The conformations of the residues are color coded according to Fig. 1(c) (i.e., α = purple, β = green, and Lα = blue for all residues except the N-terminal residue). The three depicted conformations at the start and end points of each process correspond to the three microstates to which the associated eigenvector assigns the most negative and the most positive values, respectively. In contrast to the simulations of the capped amino acids, in which all five force fields yield similar dynamic processes, the dynamic processes of AVAVA differ considerably across force fields. Note that this variation in AVAVA cannot be explained by a simple swapping or mixing of the second and third eigenvectors of the transition matrix, as shown in Figs. 4(b) and 12 of the supplementary material.[50]

The slowest dynamic process of A_{10} in the five force fields is shown in Fig. 5. For this system, we find an even wider variety of dynamic processes than in AVAVA. Most notably in the force field which are known to over-stabilize the α-helical conformation (ff03 and CHARMM27), the disruption of the α-helix, i.e., the helix-coil-transition is the slowest process. The slowest process captured by OPLS-AA/L is also the disruption of predominantly α-helical structure. In GROMOS43a1, the second slowest dynamic process can be classified as a helix-coil transition (Fig. 13 in the supplementary material[50]). However, neither in ff99SB-ILDN nor in GROMOS43a1, the slowest dynamic process involves extended regions of α-helical residue conformations.

The rates associated to the two slowest dynamic processes in AVAVA and A_{10} are shown in Table II. As for the capped amino acids, the rates vary over a large range. Since the underlying processes differ, there is no one-to-one correspondence between the rates. However, the rates can be interpreted as order parameters for the speed at which a force field explores conformational space, which is accessible with this particular force field. Note that the rates are not only influenced by the barrier heights in the system, but also by the size of the accessible conformational space, and by the shape and relative depth of the minima and thus report on the global characteristics of a force field. Interestingly, although no consistent tendency of faster versus slower exploration of the conformational space could be detected on the level of individual residues, such a gap opens up for the two larger systems. The rates for ff03 and GROMOS43a1 are consistently higher by a factor of three to ten than the rates of the other three force fields. The accessible conformational space of ff03 for AVAVA and A_{10} seems to be particularly small (Figs. 4(a) and 5(a) in the supplementary material[50]), which could explain the fast sampling. By contrast, the GROMOS43a1 simulations visited considerably more microstates than all the other force fields (Figs. 4(a) and 5(a) in the supplementary material[50]) corresponding to a very large accessible conformational space. This finding,

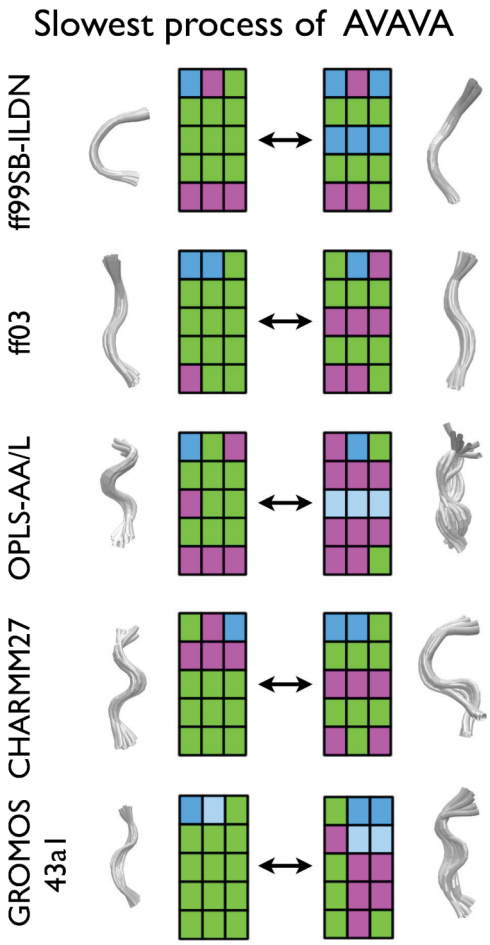

FIG. 4. Slowest dynamic process for AVAVA. For all force fields, the second eigenvector is represented via conformations between which the transition occurs. Such conformations are expressed via groups of strings color-coded as in Fig. 1, representing the three microstates to which the eigenvectors assigns the most negative and most positive values. Bundles of conformational representatives for each set of strings are provided.

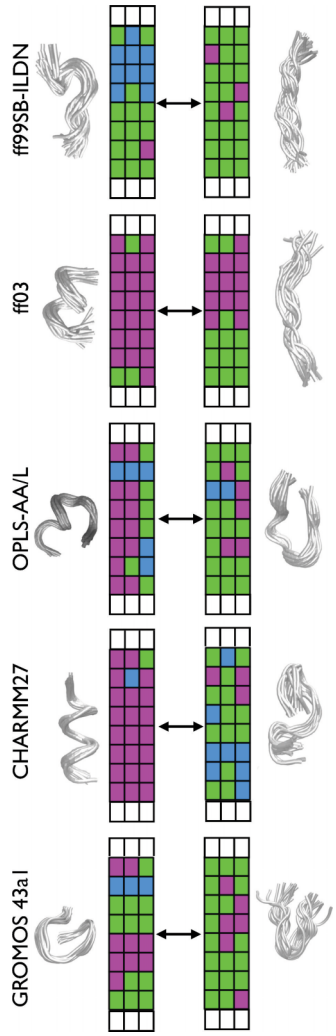

FIG. 5. Slowest dynamic process for A_{10}. For all force fields, the second eigenvector is represented via conformations between which the transition occurs. Such conformations are expressed via groups of strings color-coded as in Fig. 1 representing the three microstates to which the eigenvectors assign the most negative and most positive values. Bundles of conformational representatives for each set of strings are provided.

in combination with the high rates, hints at particularly low barriers in this force field.

Experimentally, short alanine-rich peptides are known to form helices in solutions[59–61] and therefore such systems provide an ideal basis for studying helix coil-transitions, as was recently done for a short penta-alanine,[5,62,63] as well as other systems.[64] From our MSMs, we do not expect a pronounced helix-coil transition for the studied pentapetide. Nonetheless, we did carry out an analysis of the autocorrelation times of Cα-RMSD trajectories of non-terminal residues with respect to a reference helix. For details regarding the definition of the reference structure, refer to Sec. III C. In Fig. 6(a), we show the logarithm of the ACF for all force fields as well as the logarithm of the double exponential fits to the ACFs. A single exponential model did not result in good estimates for the decay of the ACF and was therefore not used. From the fit, a set of decay times was estimated and indicated in Table II. Comparing these times to the computed relaxation times of the MSMs for each forcefield, no similarity is found. This suggests that an ACF of the RMSD trajectories, or other order parameter related to the helix-coil transition, does not give a good kinetic description of this system, as a predominate α-structure does not give rise to slow kinetics in the system. The MSM, however, allows to identify states with a high contribution to the slow dynamics without any prior knowledge of the system, e.g., irrespective of the system exhibiting a helix-coil transition or not.

On the contrary, the helix-coil transition for the deca peptide A_{10} is thought of as a valid picture for the dynamics for the system,[5,65,67] as especially the eight non-terminal residues can be found in an ideal-helix conformation multiple times throughout simulation trajectories. This helix-coil transition is thought of as a two step kinetic process, where a single exponential decay is an appropriate kinetic model. But also in the case of A_{10} it seems that a double exponential fit is a much better fit model than a single exponential. This is visualized in Fig. 6(b), where the ACF analysis again shows that a double exponential behavior describes the data best with the exception of CHARMM27, where a single exponential fit represents the data better. From Fig. 5, it is clear that the slowest process in ff03 and CHARMM27 involve a helix-coil like transition. For ff03, the fitted double exponential results in a timescale that

TABLE II. Relaxation rates k_i, the relaxation timescales t_i of the second ($i = 2$) and third ($i = 3$) dynamic process, and the fitting parameters τ_1 and τ_2 for the ACF of the simulated polymer peptide systems. For relaxation rates and timescales, we report mean and standard deviation from the bootstrap analysis.

System: AVAVA						
	Kinetic processes				ACF	
Force field	κ_2 (ns^{-1})	t_2 (ns)	κ_3 (ns^{-1})	t_3 (ns)	τ_1 (ns)	τ_2 (ns)
ff99SB-ILDN	0.031 ± 0.005	33.428 ± 5.072	0.094 ± 0.024	10.661 ± 2.272	2.18	0.29
ff03	0.228 ± 0.017	4.414 ± 0.322	0.286 ± 0.029	3.532 ± 0.352	5.49e+8	4.51
OPLS-AA/L	0.018 ± 0.008	55.481 ± 14.567	0.067 ± 0.173	14.809 ± 5.368	0.75	20.7
CHARMM27	0.065 ± 0.006	15.303 ± 1.362	0.112 ± 0.008	8.929 ± 0.659	6.71	0.66
GROMOS43a1	0.786 ± 0.072	1.273 ± 0.124	1.468 ± 0.055	0.681 ± 0.026	8.2e+2	0.65
System: A_{10}						
	Kinetic processes				ACF	
Force field	κ_2 (ns^{-1})	t_2 (ns)	κ_3 (ns^{-1})	t_3 (ns)	τ_1 (ns)	τ_2 (ns)
ff99SB-ILDN	0.021 ± 0.003	48.617 ± 6.410	0.028 ± 0.003	36.080 ± 4.466	0.88	7.61
ff03	0.116 ± 0.006	8.591 ± 0.425	0.176 ± 0.017	5.696 ± 0.539	8.60	0.78
OPLS-AA/L	0.045 ± 0.011	22.450 ± 3.941	0.068 ± 0.011	14.778 ± 2.003	1.20	7.87
CHARMM27	0.013 ± 0.005	75.664 ± 2.090	0.022 ± 0.004	44.567 ± 9.004	20.20	...
GROMOS43a1	0.118 ± 0.033	8.509 ± 2.050	0.196 ± 0.022	5.091 ± 0.629	3.09	0.43

is very comparable to the slowest timescale of the respective MSM, i.e., ~8.6 ns; for CHARMM27, this is not the case, suggesting that other configurations are responsible for a slow escape rate. In all other forcefields, the slowest timescales do not seem to be dominated by an escape from an α-helical state, suggesting that computing the ACF of the C-α RMSD trajectories does not provide a good insight into the slow kinetics of the system. The MSM in this case has the advantage that no prior information about what configurations contribute to the slow dynamics is hard coded into the analysis. In theory, the helix-coil transition may be a dominant factor in peptide dynamics, but this is not representative for all the force fields studied here.

C. Discussion

The local dynamics of capped amino acids is in qualitative agreement in all of the five tested force fields. Besides the equilibrium distribution (process I), two dominant dynamic processes were identified: (i) process II, which mediates the dynamic exchange between conformations with negative ϕ-torsion angle values and conformations with positive ϕ-torsion angle values, and (ii) process III, which mediates the dynamic exchange between the α-helical conformation and the β-sheet conformation. By contrast, the relaxation rates associated to the dynamic processes vary drastically between force fields. Moreover, the ordering of the two processes is not preserved. For most force fields, process II is the slowest process of the system, while in some cases process III is slower than process II. Our results also reproduced the known differences in the sampling of the ϕ-ψ-plane across the force fields.

The similarity of the dynamic processes at the level of single residues does not extend to the dynamic processes of small peptides. For AVAVA, all five force fields identify the formation/break-down of an elongated conformation (with three or more sequential residues in the β-sheet conformation) as the slowest process (left column Fig. 4). However, they disagree on the end point of this process (right column Fig. 4). For deca-alanine, the formation/disruption of the α-helix or a helix-like conformation was identified as the slowest process for three out of five force fields, while the other two yielded to processes which switch between structures with a large content of β-sheet conformations in the residues. The two force fields which showed the clearest helix-coil transitions—ff03 and CHARMM27—are also force fields which are known to overstabilize the α-helical structure.

Because the kinetic processes of the peptide systems differ drastically across force fields, a one-to-one comparison of the associated rates is not possible. The rates can, however, be interpreted as a measure for the overall "sampling speed" of a particular force field. While no consistent difference in the overall "sampling speed" of the force fields could be identified for the capped amino acids, two force fields (ff03 and GROMOS43a1) sampled the conformational space of the small peptide systems considerably faster than the other three force fields in this study (ff99SB-ILDN, OPLS-AA/L, CHARMM27). The fast rates of the GROMOS43a1 force field are consistent with the short relaxation timescales of capped alanine. It is however not obvious, why the rates of ff03 are higher than those of other force fields.

These results demonstrate that, even for short aliphatic peptides, the conformational dynamics cannot be reduced to a simple combination of the local dynamics of the residues. Rather the dynamic processes are shaped by a complex interaction of local dynamics and long-range interactions, including the interactions with the solvent. It is therefore not obvious which force field parameters need to be tuned to improve the dynamic performance of classical force fields. So far, reparametrizations have been focused on the bonded interactions, which are typically modeled by harmonic potentials

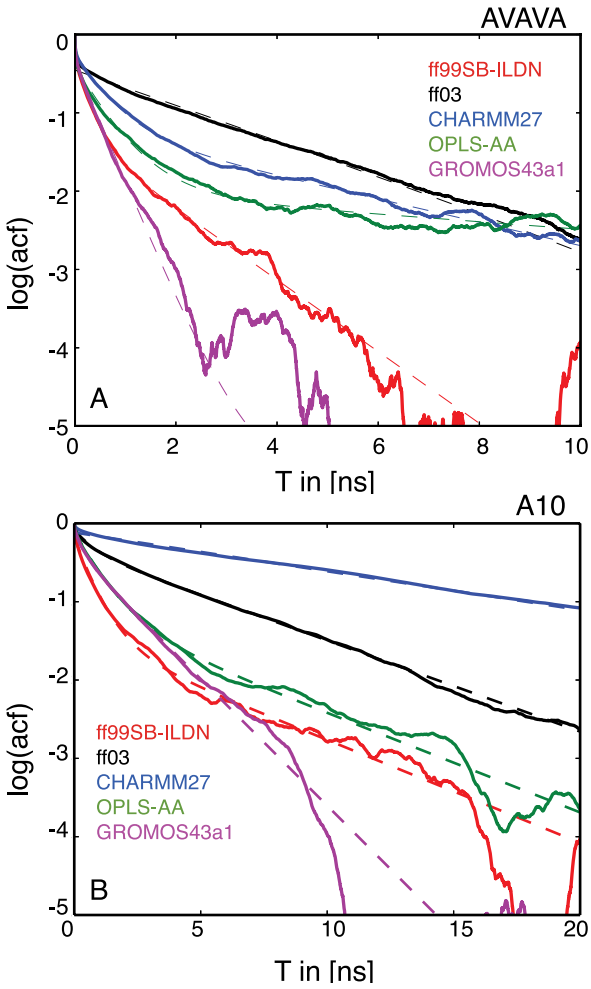

FIG. 6. Logarithm of the autocorrelation function of Cα-RMSD with respect to helical structure for AVAVA (a) and A_{10} (b). The logarithm of a double-exponential fit is shown as a dashed line. Exception is CHARMM27 for A_{10} where a single-exponential fit was used.

or comparably simple functions of the molecular coordinates. It is an open debate whether the accuracy of current force fields is limited by their simple functional form (rather than the parameters),[67,68] or whether the functional form still leaves room for improvement, e.g., by introducing amino acid dependent torsion angle potential.[16] Apart from improving the bonded interactions, a complete reparametrization of the non-bonded force field terms against state-of-the-art quantum-mechanical calculations has been suggested.[16] Last but not least, the neglect of polarizability might be a bottleneck for the ability to model dynamics. Polarization affects the stability of hydrogen bonds which in turn determine to a large part the stability of secondary structure elements. This is in line with a recent study in which the different performances of non-polarizable force fields in reproducing the mean rupture forces for the separation of mechanically interlocked dimers could be attributed to differences in modeling of the hydrogen bonds and to the resulting differences in the strength of hydrogen networks.[69]

Ultimately, to improve the dynamic properties of force fields, one will have to include kinetic experimental data into the optimization scheme. Such an optimization scheme would have to address known critical points, such as availability and choice of the experimental reference data, avoidance overfitting and cancellation of errors, and consistency with previous optimizations. But on top of this, several new challenges need to be addressed: (i) match between simulation data and kinetic experimental observables; (ii) efficient variation and validation of the force field parameters; and (iii) verification of the folding pathways predicted by the simulation.

We cannot offer a solution to these challenges at this point. However, there are several recent methodological developments which might prove to be helpful in the development of a working optimization scheme which incorporates kinetic experimental data. Issue (i) could be addressed quantitatively by dynamical fingerprints[40,41] or by simulating spectra and/or relaxation rates using a pre-parametrized Markov state model. The match between experimental kinetic data and MD simulations has been demonstrated for fluorescence quenching,[40] neutron scattering,[70,71] 2D-IR spectra,[72] as well as NMR relaxation times.[73]

To optimize the parameters of a force field towards this type of experimental data, a separate Markov state model needs to be constructed for each parameter combination. The computational cost of parametrizing each of these Markov state models from an independent simulation is prohibitive. Instead reweighting schemes, with which simulation data at one thermodynamics state point can be utilized (reweighted) to parametrize a MSM at another thermodynamic state point, could offer a solution to issue (ii). So far, reweighting has been used to make optimal use of parallel-tempering data.[74,75] An extension of the method to the variation of force field parameters is possible, thus allowing one to estimate experimental relaxation rates for a range of force field parametrizations from a single simulation. The search for optimized parameters would be further simplified if the gradient of the kinetic observable with respect to the force field parameters could be calculated. For thermodynamic data, a local minimization scheme based on explicitly calculated gradients has been published recently.[76]

The difficulty in verifying the folding pathways (issue iii) lies in extracting the folding pathway from the experimental data. Here, hidden Markov models (HMM) which are parametrized directly from single-molecule experiments,[77–80] such as single-molecule fluorescence or optical tweezer experiments, might prove useful. In a HMM analysis, the states of the MSM are defined as a function of the experimental observable (FRET efficiency, force, etc.). Provided that a structural interpretation can be found for these states, the MSM directly yields the dominant folding pathway plus all alternative folding pathways with their relative statistical weights. With this information, one can calculate the likelihood of a particular folding pathway observed in a MD simulation, thus, allowing to confer the large variation in the folding pathways across different force fields[32] with experimental evidence.

In conclusion, if MD simulations are to be used successfully in the investigation of the conformational dynamics of biomolecules, the dynamic properties of the force fields need to be taken explicitly into account in the force field development. MSMs provide as useful framework for a quantitative and detailed comparison of different force field parametrizations and to match kinetic experimental observables to MD simulations.

ACKNOWLEDGMENTS

F.V. acknowledges funding by DFG SFB 958. A.S.J.S.M. acknowledges funding by DFG SFB 740. F.N. acknowledges funding by ERC starting grant "pcCell." B.K. acknowledges funding by DFG SFB 765.

[1]Y. Duan and P. A. Kollman, "Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution," Science **282**, 740 (1998).
[2]R. H. Swendsen and J.-S. Wang, "Replica Monte Carlo simulation of spin-glasses," Phys. Rev. Lett. **57**, 2607–2609 (1986).
[3]Y. Sugita and Y. Okamoto, "Replica exchange molecular dynamics method for protein folding simulation," Chem. Phys. Lett. **314**, 141–151 (1999).
[4]J. Graf, P. H. Nguyen, G. Stock, and H. Schwalbe, "Structure and dynamics of the homologous series of alanine peptides: A joint molecular dynamics/NMR study," J. Am. Chem. Soc. **129**, 1179–1189 (2007).
[5]R. B. Best, N.-V. Buchete, and G. Hummer, "Are current molecular dynamics force fields too helical?," Biophys. J. **95**, L07–L09 (2008).
[6]F. Jiang, W. Han, and Y.-D. Wu, "Influence of side chain conformations on local conformational features of amino acids and implication for force field development," J. Phys. Chem. B. **114**, 5840–5850 (2010).
[7]F. Jiang, W. Han, and Y.-D. Wu, "The intrinsic conformational features of amino acids from a protein coil library and their applications in force field development," Phys. Chem. Chem. Phys. **15**, 3413–3428 (2013).
[8]E. A. Cino, W.-Y. Choy, and M. Karttunen, "Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations," J. Chem. Theory Comput. **8**, 2725–2740 (2012).
[9]S. Gnanakaran and A. E. García, "Helix-coil transition of alanine peptides in water: Force field dependence on the folded and unfolded structures," Proteins: Struct., Funct., Genet. **59**, 773–782 (2005).
[10]R. B. Best and G. Hummer, "Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides," J. Phys. Chem. B **113**, 9004–9015 (2009).
[11]G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides," J. Phys. Chem. B **2**, 6474–6487 (2001).
[12]V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters," Proteins: Struct., Funct., Genet. **65**, 712–725 (2006).
[13]A. D. MacKerell, N. Banavali, and N. Foloppe, "Development and current status of the CHARMM force field for nucleic acids," Biopolymers **56**, 257–265 (2001).
[14]A. D. MacKerell, M. Feig, and C. L. Brooks, "Improved treatment of the protein backbone in empirical force fields," J. Am. Chem. Soc. **126**, 698–699 (2004).
[15]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: Struct., Funct., Bioinf. **78**, 1950–1958 (2010).
[16]K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, "Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements," J. Chem. Theory Comput. **8**, 1409–1414 (2012).
[17]R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. Mackerell, "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ (1) and χ (2) dihedral angles," J. Chem. Theory Comput. **8**, 3257–3273 (2012).
[18]C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6," J. Comput. Chem. **25**, 1656–1676 (2004).
[19]O. F. Lange, D. van der Spoel, and B. L. de Groot, "Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data," Biophys. J. **99**, 647–655 (2010).
[20]K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data," PloS One **7**, e32131 (2012).
[21]J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function," Curr. Opin. Struct. Biol. **19**, 120–127 (2009).
[22]D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, "Atomic-level characterization of the structural dynamics of proteins," Science **330**, 341–346 (2010).
[23]F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, "Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations," Proc. Natl. Acad. Sci. U. S. A. **106**, 19011–19016 (2009).
[24]T. J. Lane, G. R. Bowman, K. Beauchamp, V. A. Voelz, and V. S. Pande, "Markov state model reveals folding and functional dynamics in ultra-long MD trajectories," J. Am. Chem. Soc. **133**, 18413–18419 (2011).
[25]K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," Science **334**, 517–520 (2011).
[26]K. A. Beauchamp, R. McGibbon, Y.-S. Lin, and V. S. Pande, "Simple few-state models reveal hidden complexity in protein folding," Proc. Natl. Acad. Sci. U. S. A. **109**, 17807–17813 (2012).
[27]A. K. Faradjian and R. Elber, "Computing time scales from reaction coordinates by milestoning," J. Chem. Phys. **120**, 10880–10889 (2004).
[28]P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, "Transition path sampling: Throwing ropes over rough mountain passes, in the dark," Annu. Rev. Phys. Chem. **53**, 291–318 (2002).
[29]W. F. van Gunsteren, R. Bürgi, C. Peter, and X. Daura, "The key to solving the protein-folding problem lies in an accurate description of the denatured state," Angew. Chem., Int. Ed. Engl. **40**, 351–355 (2001).
[30]P. L. Freddolino, S. Park, B. Roux, and K. Schulten, "Force field bias in protein folding simulations," Biophys. J. **96**, 3772–3780 (2009).
[31]B. Keller, Z. Gattin, and W. F. van Gunsteren, "What stabilizes the 3(14)-helix in beta3-peptides? A conformational analysis using molecular simulation," Proteins: Struct., Funct., Bioinf. **78**, 1677–1690 (2010).
[32]S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "How robust are protein folding simulations with respect to force field parameterization?," Biophys. J. **100**, L47–L49 (2011).
[33]W. C. Swope, J. W. Pitera, and F. Suits, "Describing protein folding kinetics by molecular dynamics simulations. 1. Theory," J. Phys. Chem. B **108**, 6571–6581 (2004).
[34]J. D. Chodera, N. Singhal, V. S. Pande, K. Dill, and W. C. Swope, "Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics," J. Chem. Phys. **126**, 155101 (2007).
[35]F. Noé, I. Horenko, C. Schütte, and J. C. Smith, "Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states," J. Chem. Phys **126**, 155102 (2007).
[36]N.-V. Buchete and G. Hummer, "Coarse master equations for peptide folding dynamics," J. Phys. Chem. B **112**, 6057–6069 (2008).
[37]J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: Generation and validation," J. Chem. Phys. **134**, 174105 (2011).
[38]B. Keller, P. Hünenberger, and W. F. van Gunsteren, "An analysis of the validity of Markov state models for emulating the dynamics of classical molecular systems and ensembles," J. Chem. Theory Comput. **7**, 1032–1044 (2011).
[39]J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," Curr. Opin. Struct. Biol. **25**, 135–144 (2014).
[40]F. Noé, S. Doose, I. Daidone, M. Löllmann, M. Sauer, J. D. Chodera, and J. C. Smith, "Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments," Proc. Natl. Acad. Sci. U. S. A. **108**, 4822–4827 (2011).
[41]B. G. Keller, J.-H. Prinz, and F. Noé, "Markov models and dynamical fingerprints: Unraveling the complexity of molecular kinetics," Chem. Phys. **396**, 92–107 (2012).
[42]J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, "Longtime protein folding dynamics from shorttime molecular dynamics simulations," Multiscale Model. Simul. **5**, 1214–1226 (2006).
[43]D. S. Chekmarev, T. Ishida, and R. M. Levy, "Long-time conformational transitions of alanine dipeptide in aqueous solution: Continuous and discrete-state kinetic models," J. Phys. Chem. B **108**, 19487–19495 (2004).
[44]P. G. Bolhuis, C. Dellago, and D. Chandler, "Reaction coordinates of biomolecular isomerization," Proc. Natl. Acad. Sci. U. S. A. **97**, 5877–5882 (2000).
[45]C. A. F. de Oliveira, D. Hamelberg, and J. A. McCammon, "Estimating kinetic rates from accelerated molecular dynamics simulations: Alanine dipeptide in explicit solvent as a case study," J. Chem. Phys **127**, 175105 (2007).

[46]W.-N. Du, K. A. Marino, and P. G. Bolhuis, "Multiple state transition interface sampling of alanine dipeptide in explicit solvent," J. Phys. Chem. **135**, 145102 (2011).
[47]X. Daura, A. E. Mark, and W. F. van Gunsteren, "Parametrization of aliphatic CH n united atoms of GROMOS96 force field," J. Comput. Chem. **19**, 535–547 (1998).
[48]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).
[49]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).
[50]See supplementary material at http://dx.doi.org/10.1063/1.4909549 for model validation and further analysis.
[51]M. Sarich, F. Noé, and C. Schütte, "On the approximation quality of Markov state models," Multiscale Model. Simul. **8**, 1154–1177 (2010).
[52]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**, 1701–1718 (2005).
[53]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).
[54]W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," J. Am. Chem. Soc. **7863**, 11225–11236 (1996).
[55]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).
[56]J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore, "Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution," Nature **185**, 422–427 (1960).
[57]D. S. Kosov and G. Stock, "Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments," J. Phys. Chem. B **107**, 5064–5073 (2003).
[58]T. Yoda, Y. Sugita, and Y. Okamoto, "Comparisons of force fields for proteins by generalized-ensemble simulations," Chem. Phys. Lett. **386**, 460–467 (2004).
[59]S. Marqusee, V. H. Robbins, and R. L. Baldwin, "Unusually stable helix formation in short alanine-based peptides," Proc. Natl. Acad. Sci. U. S. A. **86**, 5286–5290 (1989).
[60]S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer, "Fast events in protein folding: Helix melting and formation in a small peptide," Biochemistry **35**, 691–697 (1996).
[61]P. A. Thompson, V. Muñoz, G. S. Jas, E. R. Henry, W. A. Eaton, and J. Hofrichter, "The helix-coil kinetics of a heteropeptide," J. Phys. Chem. B. **104**, 378–389 (2000).
[62]G. Hummer, A. García, and S. Garde, "Conformational diffusion and helix formation kinetics," Phys. Rev. Lett. **85**, 2637–2640 (2000).
[63]G. Hummer, A. E. García, and S. Garde, "Helix nucleation kinetics from molecular simulations in explicit solvent," Proteins: Struct., Funct., Genet. **42**, 77–84 (2001).
[64]G. S. Jas, C. R. Middaugh, and K. Kuczera, "Non-exponential kinetics and a complete folding pathway of an α-helical heteropeptide: Direct observation and comprehensive molecular dynamics," J. Phys. Chem. B. **118**, 639–647 (2014).
[65]U. H. E. Hansmann and Y. Okamoto, "Finite-size scaling of helix-coil transitions in poly-alanine studied by multicanonical simulations," J. Chem. Phys. **110**, 1267 (1999); e-print arXiv:9810358[cond-mat].
[66]A. Hazel, C. Chipot, and J. C. Gumbart, "Thermodynamics of deca-alanine folding in water," J. Chem. Theory Comput. **10**, 2836–2844 (2014).
[67]A. D. J. Mackerell, "Empirical force fields for biological macromolecules: Overview and issues," J. Comput. Chem. **25**, 1584–1604 (2004).
[68]J. Vymetal and J. Vondrasek, "Critical assessment of current force fields. Short peptide test case," J. Chem. Theory Comput. **9**, 441–451 (2013).
[69]T. Schlesier and G. Diezemann, "Performance of different force fields in force probe simulations," J. Phys. Chem. B **117**, 1862–1871 (2013).
[70]B. Lindner, Z. Yi, J.-H. Prinz, J. C. Smith, and F. Noé, "Dynamic neutron scattering from conformational dynamics. I. Theory and Markov models," J. Chem. Phys. **139**, 175101 (2013).
[71]Z. Yi, B. Lindner, J.-H. Prinz, F. Noé, and J. C. Smith, "Dynamic neutron scattering from conformational dynamics. II. Application using molecular dynamics simulation and Markov modeling," J. Chem. Phys. **139**, 175102 (2013).
[72]W. Zhuang, R. Z. Cui, D.-A. Silva, and X. Huang, "Simulating the T-jump-triggered unfolding dynamics of trpzip2 peptide and its time-resolved IR and two-dimensional IR signals using the Markov state model approach," J. Phys. Chem. B **115**, 5415–5424 (2011).
[73]J. Xia, N.-j. Deng, and R. M. Levy, "NMR relaxation in proteins with fast internal motions and slow conformational exchange: Model-free framework and Markov state simulations," J. Phys. Chem. B **117**, 6625–6634 (2013).
[74]J.-H. Prinz, J. D. Chodera, V. S. Pande, W. C. Swope, J. C. Smith, and F. Noé, "Optimal use of data in parallel tempering simulations for the construction of discrete-state Markov models of biomolecular dynamics," J. Chem. Phys. **134**, 244108 (2011).
[75]J. D. Chodera, W. C. Swope, F. Noé, J.-H. Prinz, M. R. Shirts, and V. S. Pande, "Dynamical reweighting: Improved estimates of dynamical properties from simulations at multiple temperatures," J. Chem. Phys. **134**, 244107 (2011).
[76]M. Di Pierro, M. L. Mugnai, and R. Elber, "Optimizing potentials for a liquid mixture: A new force field for a tert-butanol and water solution," J. Chem. Phys. B **119**, 836 (2014).
[77]M. Pirchi, G. Ziv, I. Riven, S. S. Cohen, N. Zohar, Y. Barak, and G. Haran, "Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein," Nat. Commun. **2**, 493 (2011).
[78]B. G. Keller, A. Kobitski, A. Jäschke, G. U. Nienhaus, and F. Noé, "Complex RNA folding kinetics revealed by single-molecule FRET and hidden Markov models," J. Am. Chem. Soc. **136**, 4534–4543 (2014).
[79]J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gebhardt, and M. Rief, "The complex folding network of single calmodulin molecules," Science **334**, 512–516 (2011).
[80]F. Noé, H. Wu, J.-H. Prinz, and N. Plattner, "Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules," J. Chem. Phys. **139**, 184114 (2013).