JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

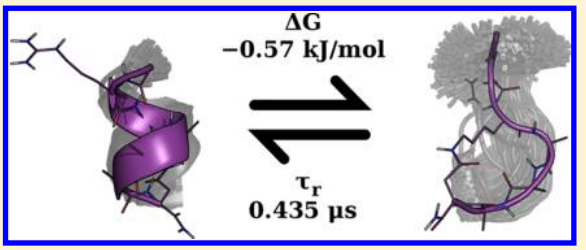
Predicting the Thermodynamics and Kinetics of Helix Formation in a Cyclic Peptide Model

João M. Damas, Luís C.S. Filipe,‡ Sara R.R. Campos,‡ Diana Lousa, Bruno L. Victor,† António M. Baptista,* and Cláudio M. Soares*

Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal

Supporting Information

ABSTRACT: The peptide Ac-(*cyclo*-2,6)-R[KAAAD]-NH_2 (*cyc*-RKAAAD) is a short cyclic peptide known to adopt a remarkably stable single turn α-helix in water. Due to its simplicity and the availability of thermodynamic and kinetic experimental data, *cyc*-RKAAAD poses as an ideal model for evaluating the aptness of current molecular dynamics (MD) simulation methodologies to accurately sample conformations that reproduce experimentally observed properties. In this work, we extensively sample the conformational space of *cyc*-RKAAAD using microsecond-timescale MD simulations. We characterize the peptide conformational preferences in terms of secondary structure propensities and, using Cartesian-coordinate principal component analysis (cPCA), construct its free energy landscape, thus obtaining a detailed weighted discrimination between the helical and nonhelical subensembles. The cPCA state discrimination, together with a Markov model built from it, allowed us to estimate the free energy of unfolding (−0.57 kJ/mol) and the relaxation time (~0.435 μs) at 298.15 K, which are in excellent agreement with the experimentally reported values (−0.22 kJ/mol and 0.42 μs, Serrano, A. L.; Tucker, M. J.; Gai, F. *J. Phys. Chem. B*, **2011**, 115, 7472−7478.). Additionally, we present simulations conducted using two enhanced sampling methods: replica-exchange molecular dynamics (REMD) and bias-exchange metadynamics (BE-MetaD). We compare the free energy landscape obtained by these two methods with the results from MD simulations and discuss the sampling and computational gains achieved. Overall, the results obtained attest to the suitability of modern simulation methods to explore the conformational behavior of peptide systems with a high level of realism.



1. INTRODUCTION

Over the past decades, atomistic molecular dynamics (MD) simulations have become an important and pervasive tool to investigate thermodynamic and kinetic properties of molecular systems and to explore the conformational space of biomolecules, particularly proteins and peptides.[1−6] These computational methods have played, and will certainly continue to play, an important role in the interpretation and rationalization of several molecular phenomena.[3,7−9] There are two major factors influencing the reliability of MD simulation studies: the sampling efficiency and the form and parameters of the potential energy function (force field).[10−14]

The ability to sample the conformational space is limited by the available computational resources. In recent years, the field of MD simulations has watched a remarkable increase in the simulation times due to a boost in the computational power and to the development of highly efficient parallelization algorithms.[10,11,15−18] Conducting MD simulations in the microsecond time range has become common practice and, in some cases, even millisecond simulations have been reported.[19−21] In addition, new methods have been developed to accelerate the sampling of the conformational space, such as replica-exchange molecular dynamics (REMD)[22] or metadynamics,[23] among many others.[3,12] Therefore, although MD simulations are still not able to cover the entire conformational space of large proteins, they are now able to exhaustively explore it for small peptides.[10−12,18]

Alongside these developments, major efforts have also been made to improve the force fields.[4,11] Although ideally a force field should accurately capture the correct balance in terms of secondary structure propensities, this is no trivial task, as demonstrated by the recent reparameterizations of several well-established force fields.[24−31]

Besides being able to capture the thermodynamic properties of proteins, force fields should also be able to provide accurate kinetic rates of relevant biological processes, such as protein folding. This is a huge challenge because most force fields are parametrized only against experimental thermodynamic and structural data.[10,13] A particularly challenging task has been the accurate description of helix nucleation events and helix−coil transition rates.[30−37] Popular force fields such as several flavors of GROMOS, AMBER, or CHARMM have been found to have some inaccuracies on helix thermodynamics and kinetics, and some improved parameters are already available.[24,27,30−32,38] The first articles comparing experimental structural and kinetic data with quantities computed from simulations employing these improved force fields have recently been pub-

Received: June 21, 2013
Published: September 19, 2013

lished.[21,39−47] A thorough evaluation through the combined usage of microsecond-timescale MD simulations and newly developed force field parameters to adequately describe these processes has been an ongoing work. Comparison with experimental data is mandatory, thus ensuring that not only an adequately weighted sampling of the helical and nonhelical subensembles is being obtained but also a correct kinetic description of the transitions between those subensembles. An obvious approach is to use simple model systems (normally short peptides) with a small number of degrees of freedom and extensive experimental characterization.

In the present work, we study the peptide Ac-(*cyclo*-2,6)-R[KAAAD]-NH_2 (hereafter referred to as *cyc*-RKAAAD, Figure 1), a cyclic hexapeptide known to adopt a remarkably stable α-

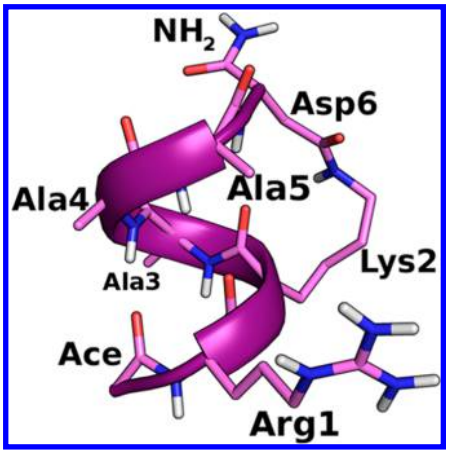


Figure 1. Representation of the *cyc*-RKAAAD peptide in an α-helical conformation.

helical conformation in water.[49,48] This peptide was initially devised by Fairlie and co-workers as a minimalistic model for α-helix formation.[48] The available experimental structural data show that indeed the peptide adopts an α-helical conformation, stabilized by the cyclizing linkage between side chains.[49,48] Recently, Gai and co-workers have also examined *cyc*-RKAAAD conformational relaxation kinetics using temperature-jump (T-jump) experiments.[49] The conformational simplicity of *cyc*-RKAAAD and the availability of experimental data make it an excellent model to study the ability of modern MD simulations to adequately describe the balance between helical and nonhelical states, as well as the underlying kinetics.

Herein, we extensively sample the conformational space of *cyc*-RKAAAD using microsecond-timescale MD simulations. To perform a comprehensive characterization of the peptide conformational space, we first analyze its secondary structure preferences and then use the Cartesian-coordinate principal component analysis (cPCA) to describe the peptide free energy landscape and discriminate the conformational states, after which we employ a Markov model formalism to estimate the kinetic relaxation time constant. Finally, we analyze the sampling obtained from standard MD simulations with the one from two enhanced sampling methods, REMD and bias-exchange metadynamics (BE-MetaD), with the objective of comparing the simulation times and computational demands needed to obtain a similar description of the *cyc*-RKAAAD conformational space. The present work also shows that the GROMOS 54A7 force field[24] is able to reproduce the thermodynamics and kinetics of helix formation.

2. METHODS

2.1. Setup for MD Simulations. The cyclic peptide *cyc*-RKAAAD was built in a helical conformation using PyMOL version 1.2r1 (Schrödinger, LLC). For the molecular mechanics description of the system, the 54A7 GROMOS force field[24] and the SPC water model[50] were used. For the cyclization between N^{ζ} of Lys2 and C^{γ} of Asp6, a standard peptide bond was considered, meaning that all bonded parameters (bond, angles, and dihedrals) involving that bond were added or modified accordingly. For the molecular dynamics simulations, the GROMACS 4.5.1 package[17] was used. The simulations were performed using periodic boundary conditions at 298.15 K and 1 atm, with an integration time step of 0.002 ps. The temperature and pressure were kept constant with Berendsen coupling baths,[51] with separate temperature coupling for the peptide and solvent. The pressure coupling constant was 0.5 ps, the isothermal compressibility was 4.5×10^{-5} bar^{-1}, and the heat coupling constant was 0.1 ps. The twin-range cutoff method[8] was applied to nonbonded interactions, with short- and long-range cutoffs of 9 and 14 Å, respectively. A reaction-field correction[52] was applied to electrostatic interactions, considering a value of 54 for the dielectric constant.[53] The neighbor lists were updated every five steps. All bonds were constrained to their equilibrium lengths with the P-LINCS algorithm,[54] except the water molecules, which were kept rigid with the SETTLE algorithm.[55] The prepared system was subjected to 50 steepest-descent steps of energy minimization to let the side-chains of Lys2 and Asp6 adapt to the cyclization. Then, in order to obtain a generic "unstructured" starting point for the simulations, the partial charges of each atom were changed from their reference values to +0.5*e*, and 1000 steepest-descent steps of minimization were performed. This brief change in all atomic electric charges promotes the repulsion among all atoms of the peptide, allowing the simulations to depart from a nonhelical conformational state.[56] This starting conformation was then solvated in a rhombic dodecahedral box, considering a minimum distance of 10 Å between the peptide and the box walls, ending up with a total of 1154 water molecules. Then, we performed three simulations of standard MD, initiated with different sets of random velocities sampled from a Maxwell−Boltzmann distribution. Each simulation was run for 10 μs, for a total of 30 μs of sampling time. Conformations were sampled every 10 ps.

2.1.1. Specific Details of REMD Simulations. REMD is an enhanced sampling method where the conformational space is explored in parallel by different replicas, each simulated at a different temperature. Replicas can attempt to exchange configurations at regular time intervals using a Metropolis criterion, enabling the system to cross high energy barriers.[22]

We have performed three REMD simulations of *cyc*-RKAAAD starting from the same conformations that were used to initiate the standard MD simulations described above. These simulations correspond to independent REMD calculations, initiated with different random velocities taken from a Maxwell−Boltzmann distribution, and they should not be confused with the temperature replicas that are used in each one of them. The simulation setup was identical to the one used in the standard MD simulations, with the exception of the temperature coupling algorithm, which in this case was V-rescale.[57] For each individual REMD simulation, we have explored a range of temperatures between 290 and 345 K, in a

total of 12 different temperatures (290.00, 294.75, 299.55, 304.42, 309.35, 314.34, 319.39, 324.51, 329.69, 335.07, 340.39, and 345.78 K). This temperature distribution was predicted[58] to guarantee that a reasonable exchange rate between neighbor temperatures was achieved (accepting 20% of the exchange attempts). Each REMD simulation started with an equilibration period of 10 ns, followed by a productive period of 1 μs, in which exchange attempts were made with a 1 ps interval. In total, for all replicas of the three simulations, we have simulated the peptide for 36 μs.

2.1.2. Specific Details of BE-MetaD Simulations. Metadynamics is a technique which enhances sampling by accelerating rare events. In this method, the system is discouraged from revisiting previously sampled configurations by the application of a time-dependent external potential, which acts on the space of a small number of degrees of freedom called collective variables (CVs).[23,59] This enables the system to escape low energy regions and rapidly explore the conformational space. BE-MetaD is a variant of metadynamics, in which several replicas are run in parallel (like in REMD), with the bias being applied to different CVs in each replica.[60] As in REMD, exchanges between replicas are attempted from time to time and accepted or rejected with a probability given by a Metropolis criterion.

We have performed three independent BE-MetaD simulations[60] using the PLUMED plugin[61] for GROMACS 4.5.5, starting from the conformations obtained after 50 ns of standard MD. The simulation setup was identical to the one that was used for REMD. Six replicas were used in each BE-MetaD simulation. Each of the first five replicas was biased by a different CV, namely main-chain radius of gyration (R_g), mean square deviation (MSD) from *cyc*-RKAAAD in a perfectly helical structure, number of main-chain hydrogen bonds contributing to α-helical structures (N_{hb}), helicity of the main-chain (Φ_a) (with reference values of −64 and −41 degrees for phi and psi angles, respectively), and dihedral correlation (Φ_{corr}) (the definitions of these CVs can be found in ref 61). The sixth replica was not biased by any external potential (neutral replica). The replicas can attempt to exchange every 10 ps, and Gaussian potentials with a height of 0.1 kJ/mol were added every 1 ps. These values have been shown to provide a good compromise between speed and accuracy.[60] The width of the Gaussians, which determines the resolution of the free energy reconstruction,[60] was 0.005 nm, 0.0005 nm^2, 0.05, 0.05, and 0.05 for the collective variables R_g, MSD, N_{hb}, Φ_a, and Φ_{corr}, respectively. We used small Gaussian widths (considerably smaller than the smallest fluctuations observed in standard MD) to obtain a high resolution in the reconstruction of the free energy.

BE-MetaD simulations have an equilibration period, after which the free energy profiles as a function of the different CVs start to converge and the population of the neutral replicas starts to stabilize.[60] To analyze the population of the neutral replicas, we built histograms of the mean square displacement of the structures sampled in these simulations from *cyc*-RKAAAD in a perfectly helical structure. In these histograms, three clear peaks are visible, corresponding to different states of the peptide (perfect helix, distorted helix, and nonhelical). We performed tests where the first 10, 20, 30, and 40 ns of simulation were discarded, and we observed that the histograms stabilize after 20 ns, which corresponds approximately to the time at which the free energy profiles start to converge. Therefore, we considered an equilibration period of 20 ns, which was followed by a production period of 80 ns.

2.2. Conformational Analysis. Principal component analysis (PCA) is a method that transforms a set of variables by rotation of the axes, thus obtaining a new set of uncorrelated variables called principal components (PCs), which are ordered by their variance.[62] The selection of only a few of the first PCs usually yields a good set of collective variables, thus allowing us to reduce the dimensionality while retaining most of the data distribution pattern. We used PCA to represent *cyc*-RKAAAD conformational space in only two dimensions and calculate the free energy landscape following a previously published approach.[63]

PCA was performed on the Cartesian coordinates of the main-chain atoms (cPCA) after fitting to a central structure (see definition in ref 63), using a sample of all structures saved during the standard MD simulations. The first two PCs accounted for more than 80% of the total variance. The two-dimensional (2D) space thus obtained was also used to project the samples of structures from either the BE-MetaD or the REMD simulations, for comparison purposes.

A kernel estimate of the probability density[64] was computed on a grid of (0.2 Å)^2 bins, using a Gaussian kernel, with a bandwidth calculated as in ref 63. The probability density surface was then converted to a free energy surface according to:

$$F(\boldsymbol{r}) = -RT\ln\frac{P(\boldsymbol{r})}{P_{max}}$$

where $\boldsymbol{r}$ is the coordinate in the 2D space of the first PCs and P_{max} is the maximum of the probability density function, $P(\boldsymbol{r})$. The quantity $F(\boldsymbol{r})$ is a partial or conditional free energy, corresponding to a sum-overstates over all degrees of freedom except the first two PCs. The free energy landscapes were analyzed by determining the energy minima and assigning the respective basins using a steepest-descent criterion.[63] After having assigned each structure of the simulation trajectory to a basin, the transition matrix can be directly calculated.

2.3. Relaxation Time Determination from Lifetimes. The relaxation time of a two-state process can be estimated using the stable-states picture with a dual-cutoff approach previously applied to protein folding/unfolding.[20,37] After defining the folded and unfolded states (in this work, helical and nonhelical) using a simple RMSD criterion, one registers the lifetimes of each state along the simulation (i.e., the between-transitions time intervals during which the system spends in that state). Averaging the lifetimes of the folded or the unfolded state yields an estimate of the characteristic times of unfolding (τ_u) or folding (τ_f), respectively. A relaxation time estimate τ_r may then be obtained through the relation $1/\tau_r = 1/\tau_f + 1/\tau_u$. This estimation obviously improves with a higher number of folding and unfolding events and a better sampling of lifetimes, as the averaging improves accordingly.

2.4. Markov Model Construction and Iteration. If we consider the temporal evolution of the system in the state space as a stochastic process, and if we assume it to be Markovian (more on this assumption in the Results and Discussion), we may construct a Markov model to describe its dynamics.[65] In order to do so, we must identify the sampled conformations in the state space and then infer the underlying transition stochastic matrix from the trajectories transition statistics. The identification of the conformations in the state space was

derived from the cPCA analysis (see above). Taking into account the obtained landscape and the helix-formation process in which we were interested, we have considered two states, helical and nonhelical, each containing several basins (see Results and Discussion). The assignment of conformations to a state is taken directly from which basin they belong. Using the temporal jumps given by the trajectories and an overlapping window count method[65] for a given lag-time Δt, we can infer each element of the transition count matrix, $c_{ij}(\Delta t)$, which corresponds to the number of observed transitions from state i to state j in time Δt. Each element of the transition stochastic matrix $T_{ij}(\Delta t)$ is then given by

$$T_{ij}(\Delta t) = \frac{c_{ij}(\Delta t)}{\sum_k c_{ik}(\Delta t)}$$

with the matrix describing, in this case, the two-state Markov model for the dynamics of the peptide. If we denote the probability of state i at time t as $p_i(t)$, the single-transition evolution for the Markov chain is given by

$$p_j(t + \Delta t) = \sum_i p_i(t) T_{ij}(\Delta t)$$

The iteration of this Markov chain for $t \to \infty$ gives the equilibrium or stationary probability $\pi_i = p_i(\infty)$, which obeys the invariance relation $\pi_j = \Sigma_i \pi_i T_{ij}(\Delta t)$ and, since there are only two states, also the detailed balance relation $\pi_i T_{ij}(\Delta t) = \pi_j T_{ji}(\Delta t)$. The iteration process from any given starting probability distribution $p_i(0) \neq \pi_i$ corresponds to a relaxation process toward π_i, which in the case of a two-state Markov model is necessarily given by

$$p_i(t) = \pi_i - [\pi_i - p_i(0)]\exp[-t/\tau_r(\Delta t)]$$

where $\tau_r(\Delta t)$ is the relaxation time (there is a single relaxation mode in this two-state model). Thus, the value of $\tau_r(\Delta t)$ can be obtained by fitting a single-exponential decay function to the $p_i(\Delta t)$ values generated by the Markov chain iteration for any of the two states.

2.5. Overlap of Probability Densities. One way to inspect the convergence of the simulations involves calculating a measure of the overlap of the probability density functions in the 2D space after different time lengths. Such overlap measure can also be used to compare the sampling obtained by different methods when the structures are represented in the same space.

The difference between two densities $P_1(\boldsymbol{r})$ and $P_2(\boldsymbol{r})$ defined in the same d-dimensional space is often measured using the integrated square error[64]

$$\int [P_1(\boldsymbol{r}) - P_2(\boldsymbol{r})]^2 \mathrm{d}\boldsymbol{r} = \int P_1^2(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + \int P_2^2(\boldsymbol{r})\mathrm{d}\boldsymbol{r} - 2\int P_1(\boldsymbol{r})P_2(\boldsymbol{r})\mathrm{d}\boldsymbol{r}$$

where the integrals extend over the whole d-dimensional space. A convenient overlap measure can be obtained by simply rearranging the terms of the integrated square error, yielding a normalized integrated product (NIP):

$$\mathrm{NIP} = \frac{2\int P_1(\boldsymbol{r})P_2(\boldsymbol{r})\mathrm{d}\boldsymbol{r}}{\int P_1^2(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + \int P_2^2(\boldsymbol{r})\mathrm{d}\boldsymbol{r}}$$

which gives a value between 0 (no overlap) and 1 (perfect overlap). The NIP can be analytically computed when the densities are Gaussian kernel estimates (see Supporting Information).

3. RESULTS AND DISCUSSION

3.1. Overall Trends. *cyc*-RKAAAD is able to adopt a single turn α-helix conformation in water.[49,48] Using a secondary structure criterion (DSSP[66]) and a similarity measure (main-chain RMSD) against a perfect α-helix, the conformational behavior of the peptide in each standard MD simulation was followed along time (Figure 2). The peptide is indeed capable

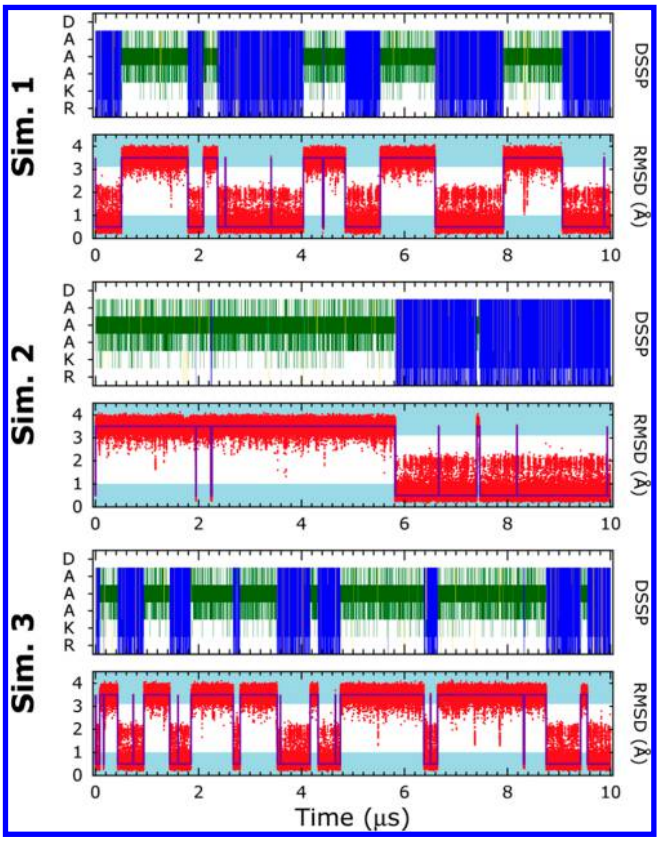


Figure 2. Time evolution of *cyc*-RKAAAD structural properties in standard MD simulations. For each of the three MD simulations, secondary structure (top) according to DSSP[66] and main-chain RMSD (bottom) computed against *cyc*-RKAAAD perfect helical structure are depicted. In the DSSP, blue, green, yellow, and white represent α-helix, bend, turn, and coil, respectively. RMSD is shown as red dots and shadowed areas delimit two states, helical (lower) and nonhelical (upper), with purple lines accounting for transitions between them.

of adopting conformations within 1 Å RMSD from a perfect α-helix, which is further corroborated by the concomitant identification of the α-helix structure with DSSP. This result is in line with determined NMR structures, which are within a main-chain RMSD of 0.81 ± 0.35 Å from a perfect α-helix.[48] Besides that, the peptide is able to adopt other, nonhelical conformations, spending much time over 3 Å RMSD from an α-helix in a bended conformation. Throughout time, the peptide is shown to interconvert between these two clear subpopulations or states (defined based on the RMSD histogram): a helical state (RMSD less than 1 Å) and a nonhelical state (RMSD more than 3.1 Å). Interestingly, even these long-timescale simulations show distinct trends, with simulation 2 spending most of the first 6 μs in the nonhelical state and most of the last 4 μs in the helical state, contrasting

dx.doi.org/10.1021/ct400529k | J. Chem. Theory Comput. 2013, 9, 5148−5157

with the more regular interconversions in simulations 1 and 3. However, using this RMSD criterion to identify the persistence of states and the transitions between them (respectively, horizontal and vertical purple lines in Figure 2), simulations 1 and 2 are the most similar pair when the lifetime-estimated relaxation times (see section 2.3) are compared (0.191, 0.202, and 0.115 μs for simulations 1, 2, and 3, respectively), while simulations 2 and 3 are the most similar pair when ratios of nonhelical/helical conformations are compared (0.84, 1.41, and 1.77 for simulations 1, 2, and 3, respectively). This shows how similarity strongly depends on the property being considered, and this diverse behavior of simulations clearly illustrates the importance of not relying on a single simulation, even for microsecond-timescale simulations. The average structural features observed in the simulations were directly compared with the available NMR data. As shown in the Supporting Information, the computed chemical shifts, *J*-couplings, and ROE violations are in very good agreement with the experimental ones. For these properties, no significant difference exists between the three simulations, perhaps reflecting similar conformational distributions.

3.2. Conformational Analysis. Having analyzed the overall trends of the individual simulations, a more detailed analysis of the type of conformations adopted and their populations was performed with the set of *cyc*-RKAAAD structures obtained from the three simulations taken together. A good description of the conformational space is given by the free energy landscape in the 2D space of the cPCA shown in Figure 3, which accounts for more than 80% of the total variance. A dihedral PCA using the sine/cosine of φ/ψ torsions was also tested, but the resulting landscapes were more rugged, as previously observed,[67] and there was not a perceptible structural distinction between the basin-specific groups of nonhelical conformations. Thus, only the cPCA was pursued.

As can be seen from Figure 3, five basins were found, corresponding to five structurally distinct free energy minima. In addition, each basin encompasses a structurally distinct and homogeneous cluster, as can be observed in the right panel of Figure 3, supporting the suitability of this 2D conformational space. The basins were labeled from 1 to 5 in ascending order of the corresponding free energy minimum. The helical conformations were clustered in basins 1 and 3, with the conformations in the latter presenting a disordered N-terminus. On the other hand, nonhelical conformations are clustered in basin 2, presenting more extended conformations, and basin 4, with more compact ones. Finally, an intermediate type of conformation, containing a persistent hydrogen bond in a position necessary for the establishment of an α-helical structure (NH proton of Ala5 and CO oxygen of Arg1), is observed in basin 5.

Accounting for more than 94% of the conformations and containing all structures with free energy below 3*RT*, basins 1 (39.53%) and 2 (54.69%) are the main clusters observed (Figure 3). While basin 1 is funneled, deep, and narrow, favoring the formation of a perfect helix, basin 2 is larger and shallower, containing a variety of extended nonhelical conformations. Overall, basin 2 presents a lower free energy due to this entropic effect, with the formation of nonhelical conformations being favored. In addition, a high number of transitions is observed between basins 1 and 3 and also between basins 2 and 4. In that way, we can think of basins 1 and 3 as a superbasin clustering the helical conformations (44.30%) and of basins 2 and 4 as another superbasin clustering the

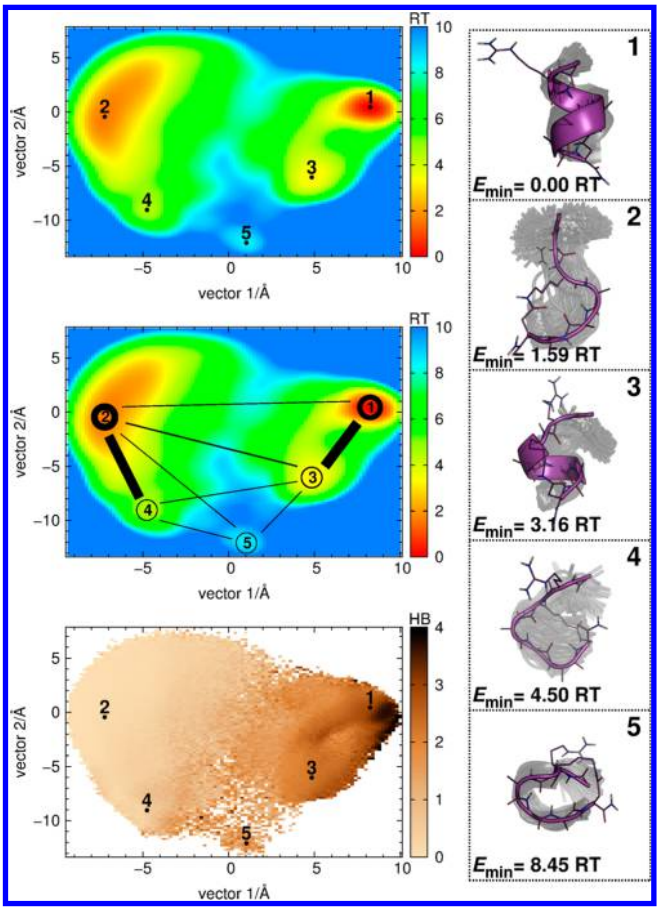


Figure 3. Free energy landscape and conformational analysis of *cyc*-RKAAAD. Free energy landscape (left, top) projected on the first two cPCA vectors obtained from standard MD data. The basins found on this landscape are labeled from 1 to 5. The upper landscape highlights the position of each free energy minima (black dots). The structures corresponding to these free energy minima (right) are represented in purple together with structures sampled from the same basin (background, gray colored) following the same labeling. The interbasin transitions are also depicted on the landscape (left, middle) with the thickness of the open-circles and lines proportional to the basin population and the number of transitions, respectively. The average number of main-chain hydrogen bonds in the 2D space is also depicted (left, bottom). The error of the free energy landscape can be found in the Supporting Information.

nonhelical structures (55.69%). In contrast, basin 5 is isolated in a high-free-energy region, accounting for barely 0.01% of the conformations, and we observe a very low number of transitions between this and all of the other basins. The transitions among all basins are schematically depicted in the left middle panel of Figure 3; notice that no transitions were observed between basin 1 and basins 4 or 5.

Furthermore, we found an interesting relationship between the free energy landscape thus obtained and the hydrogen bond content of the main chain. In the left lower panel of Figure 3, we represent the average number of intra main-chain hydrogen bonds observed at each position in the space of the first 2 PCs that was used to build the free energy landscape. We can see an overall tendency of formation of hydrogen bonds as we go from the nonhelical region (superbasin 2 + 4) toward the helical region (superbasin 1 + 3). The most frequent hydrogen bonds formed in each basin are listed in the Supporting Information. A well-defined α-turn should have four $i \rightarrow i + 4$ hydrogen

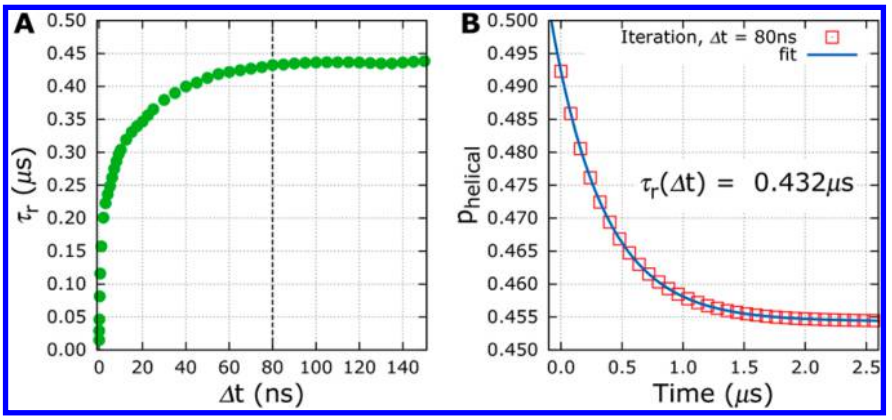


Figure 4. Calculation of *cyc*-RKAAAD relaxation time (τ_r) using a Markov chain. (A) τ_r dependence on the lag-time (Δt) used to construct the transition stochastic matrix. The τ_r for each Δt is obtained through a single-exponential fit to the variation of the helical state probability ($p_{helical}$) from a nonequilibrium distribution by iteration of the Markov chain for the two-state Markov model. This procedure is illustrated for Δt = 80 *ns* in (B).

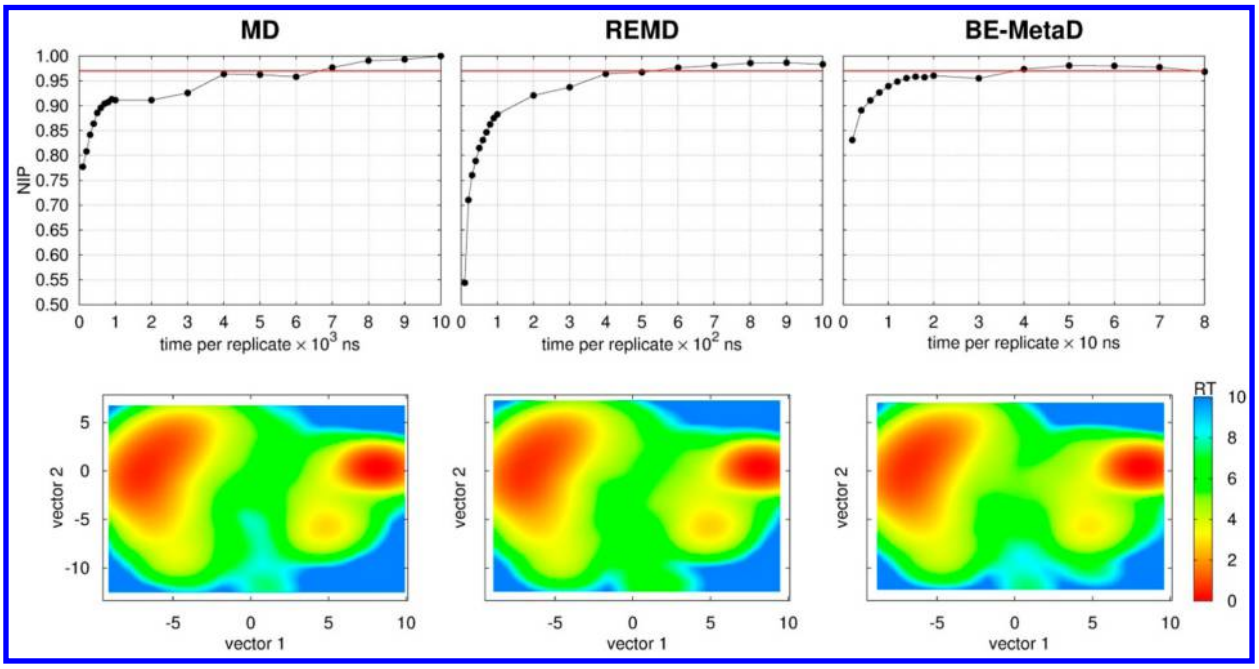


Figure 5. Comparison of different sampling strategies: standard MD (MD), replica exchange MD (REMD), and bias-exchange metadynamics (BE-MetaD). The plots in the top row show the overlap (calculated using the NIP measure, see Methods) between the landscape obtained by each method after a given simulation time and the landscape obtained using the three 10 μs simulations of standard MD (reference landscape). The red line highlights the cutoff that was used to estimate the minimum time required to reproduce the reference landscape (NIP = 0.97). We note that the time scale is different in each plot. The bottom row displays the free energy landscapes obtained by each method, using the minimum time required to reproduce the reference landscape, with the sampled structures projected in the reference cPCA space. Snapshots collected with intervals of 1000, 100, and 10 ps were used to construct the MD, REMD, and BE-MetaD landscapes, respectively. We note that the free energy zero is the same in all the landscapes.

bonds involving the NH protons of the C-terminal amide, Asp6, Ala5, and Ala4 together with, respectively, the CO oxygens of Ala3, Lys2, Arg1, and N-terminal amide. In general, basins 1 and 3 present the $i \rightarrow i + 4$ hydrogen bonds characteristic of the α-helix geometry, but basin 3 very seldom presents a hydrogen bond between Ala4 and the N-terminal amide, in agreement with the lack of structure observed at this terminus. In fact, in basin 3, Ala4 prefers to establish an $i \rightarrow i + 3$ hydrogen bond with Arg1, perhaps explaining the previously observed deviation of the exocyclic Arg residue from the canonical α-helix geometry.[48] Also, the hydrogen bond between the C-terminal amide and Ala3 is observed less frequently than the other bonds, in both basins.

Using the enthalpy and entropy for unfolding estimated by Gai and co-workers[49] from the fitting of a CD thermal curve to a two-state model, the free energy of unfolding at 298.15 K is −0.22 kJ/mol. The ratio of nonhelical/helical conformations obtained by us on the basis of cPCA separation of states (nonhelical: basins 2,4, and 5; helical: basins 1 and 3) is 1.26, which corresponds to a free energy of unfolding equal to −0.57 kJ/mol, in excellent agreement with the experimental value.

3.3. Kinetics of Helix/Non-Helix Transitions. Since the obtained free energy landscape of the peptide seems to provide a very good description of the state space, we have built a two-state Markov model according to the description. This kind of approach has become quite popular in recent years, especially

for the study of peptide and protein folding.[42,68−70] Here, we use the Markov model to better estimate the relaxation time, τ_r, of *cyc*-RKAAAD, and compare it to the experimentally obtained one. As described in the Methods section, we iterated the Markov chain starting from a nonequilibrium probability distribution and fitted a single-exponential function to the variation of the helical state probability ($p_{helical}$) along time, which allowed us to derive a value for τ_r (illustrated for Δt = 80 *ns* in Figure 4B). Nevertheless, this procedure relies on the assumption that the dynamics of the process is, indeed, approximately Markovian at a given lag-time Δt. To assess this, we checked whether a plot of $\tau_r(\Delta t)$ versus Δt converges to a constant value, as it must once Markovianity is reached.[71] The model construction and iteration process was then repeated for many lag-times, and it was clear that such a convergent behavior was present, indicating that the process is Markovian for $\Delta t \geq 80$ ns (Figure 4A). This converged τ_r is approximately 0.435 μs, which is in excellent agreement with the 0.42 μs value derived from T-jump infrared experiments.[49]

This estimated relaxation time relies on the use of transition statistics coming from three simulations. As was seen before (Figure 2), simulation 2 presents helix-transition trends different from simulations 1 and 3, and accordingly, it presents a higher relaxation time than the other two simulations when calculated from Markov models built for each simulation alone (Figure SI4). Nonetheless, the estimation for simulation 2 is compromised because for lag times where Markovianity is met, the stationary probabilities coming from the Markov model diverge from the ones coming from the MD trajectories (Figure SI5). Despite that, the limitations of simulation 2 are clearly mitigated when the global statistics from the three simulations altogether are taken into account. This underlines once more the importance of using multiple simulations, especially when taking kinetic properties into consideration.

3.4. Sampling Strategies. After showing that long MD simulations are able to capture the thermodynamic and kinetic properties of this peptide, it would be interesting to know if the free energy landscape could be reproduced using less CPU time. To answer this question, we tested two different sampling strategies against standard MD: REMD and BE-MetaD. In order to compare the sampling efficiency of the three methods, we analyzed the simulation time which is required to reproduce the reference free energy landscape, which was built using all three 10 μs simulations of standard MD. All the simulation times presented in this section are average values over the three simulations run for each method (see Methods). For REMD, we analyze only the trajectories obtained at 299.55 K, which is close to the temperature of the standard MD, while for BE-MetaD we analyze only the neutral replicas, which have an approximately canonical distribution.[60] We note that we could have used all the replicas to build the BE-MetaD energy landscape. This strategy has been shown to provide accurate results, and it is very efficient from the computational viewpoint.[72] However, in this work, we decided to follow the approach described in the first paper reporting BE-MetaD simulations,[60] where the authors focused on the analysis of the population of the neutral replica. This approach allowed us to perform the analysis presented in Figure 5, relying on a uniform treatment of the three methods (MD, REMD, and BE-MetaD) under comparison.

Figure 5 (top row) shows the overlap between the free energy landscape obtained after a given simulation time and the reference landscape. The three overlap plots have a similar profile, although REMD and BE-MetaD converge, respectively, ∼10 and ∼100 times faster than standard MD. Given that an overlap of at least 0.97 is obtained by all the methods, we chose this cutoff (red horizontal line in the plots) to estimate the time required to effectively sample the conformational space. The estimated time is 7000, 600, and 40 ns (per simulation) for standard MD, REMD, and BE-MetaD, respectively. Despite the time taken, these two alternative methods also sample conformations whose NMR properties are in good agreement with the experimental data (see Supporting Information).

Using the minimum required time estimated for each method, we built the corresponding free energy landscapes (bottom row of Figure 5). The landscapes generated using the three sampling strategies are very similar. The most populated basins are found by all the methods and have similar free energy values in all of them. The free energy barriers between the helical and nonhelical basins are slightly higher in standard MD than in REMD and especially in BE-MetaD, which probably reflects the tendency of standard MD to poorly sample rare events and consequently overestimate the population-inferred free energy of those regions. Additionally, the small differences between the reference and BE-MetaD landscapes can arise from the fact that the distribution of the neutral replica would only be rigorously equivalent to an unbiased trajectory when the Gaussian deposition rate approaches zero.[60]

We have seen that the simulation time needed to sample the conformational space using REMD and BE-MetaD is ∼1/12 and ∼1/175 of the time required by a standard MD simulation. However, these values cannot be directly used to compare the computational efficiency of the methods because, unlike standard MD, the other two methods use replicas and can have different performances. Therefore, in order to compare the three methods in a consistent way, we analyzed the performance of each method. We performed three simulations of 100 ps in an AMD Opteron Processor 6376 with 64 cores using two cores per replica. The average values obtained were 28.9 ± 0.08, 29.1 ± 0.14, and 31.6 ± 0.45 ns/day for standard MD, REMD, and BE-MetaD, respectively. Standard MD required three independent simulations of 7000 ns, which amounts to $(3 \times 7000)/28.9 \sim 727$ days of CPU time (in two cores). REMD required three independent simulations of 600 ns to sample the conformational space. However, each simulation used 12 temperature replicas, which means that the CPU time needed was $(3 \times 12 \times 600)/29.1 \sim 742$ days (in two cores). BE-MetaD required three independent simulations of only 40 ns to obtain a good sampling. In this case, we had six bias-exchange replicas per simulation, meaning that the required time was $(3 \times 6 \times 40)/31.6 \sim 23$ days of CPU (in two cores). Therefore, the computational efficiency of standard MD and REMD is identical, whereas BE-MetaD is ∼30 times more efficient than the other two methods. Nonetheless, the efficiency of the different sampling strategies will probably depend on the system under study, and therefore, extrapolations based on these results require caution.

4. CONCLUDING REMARKS

In this work, we performed microsecond-timescale MD simulations of the cyclic peptide *cyc*-RKAAAD, known to adopt a stable single-turn α-helix in water.[49,48]

From the analyses of the secondary structure propensities using DSSP and main-chain RMSD against an α-helix, two main states were observed: helical and nonhelical. A more detailed conformational analysis was performed by calculating a

free energy landscape from the probability density of the sampled structures in the 2D space of cPCA. Two low-free-energy basins were found containing either perfect helix conformations (basin 1) or extended nonhelical conformations (basin 2), in addition to three low-populated, high-free-energy basins presenting helical conformations with a disordered N-terminus (basin 3), compact nonhelical conformations (basin 4), or an intermediate type of structure (basin 5). We showed that this classification, while naturally emerging from the topography of the free energy landscape, was still capable of reflecting both the secondary structure and the hydrogen bond content, properties that can be related with experimentally measured signals. The definition of only two states (nonhelical: basins 2, 4, and 5; helical: basins 1 and 3) allowed us to estimate the free energy of unfolding and relaxation time, the latter using a two-state Markov model, in good agreement with values obtained experimentally.[49] Therefore, the force-field GROMOS 54A7 seems capable of reproducing the thermodynamics and kinetics of helical/nonhelical equilibrium in such a simple model.

In addition, the comparison of three different sampling strategies (standard MD, REMD, and BE-MetaD) showed that both REMD and BE-MetaD are able to reproduce in less time (1/12 and 1/175, respectively) the conformational distribution of a small cyclic peptide observed in long-time standard MD. While REMD is equivalent to standard MD in terms of computational efficiency, BE-MetaD is particularly efficient (~30 times saving of CPU time).

Our results illustrate how modern MD simulation protocols, combining state-of-the-art force fields with simulations on the microsecond time range, allow an excellent description of the thermodynamic and kinetic features of peptide systems.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Calculus of the normalized integral product (NIP), NMR properties calculations (chemical shifts, $^3J(H^N\text{-}H^\alpha)$ coupling constants, ROE upper distance violations), error of the free energy landscapes, hydrogen bonds analysis, and Markov models of separate simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: claudio@itqb.unl.pt (C.M.S.).
*E-mail: baptista@itqb.unl.pt (A.M.B.).

Present Address

[†]BSIM[2]—Biomolecular Simulations, Parque Tecnológico de Cantanhede, Núcleo 04, Lote 2, 3060−197 Cantanhede, Portugal.

Author Contributions

[‡]L.C.S. Filipe and S.R.R. Campos contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through Grant No. PEst-OE/EQB/LA0004/2011. J.M.D. and L.C.S.F. are also supported by FCT through Ph.D. fellowships SFRH/BD/41316/2007 and SFRH/BD/76085/2011, respectively, and B.L.V through a postdoctoral fellowship (SFRH/BPD/29708/2006). We would also like to thank Pedro R. Magalhães for the early implementation of the GROMOS54A7 force-field on GROMACS and Professor David Fairlie for assistance and providing the NMR structures from reference 48.

■ ABBREVIATIONS

cyc-RKAAAD, Ac-(*cyclo*-2,6)-R[KAAAD]-NH_2; MD, molecular dynamics; PCA, principal component analysis; cPCA, Cartesian-coordinate PCA; REMD, replica-exchange molecular dynamics; BE-MetaD, bias-exchange metadynamics; T-jump, temperature-jump; CVs, collective variables; MSD, mean square deviation; PCs, principal components; 2D, two-dimensional; NIP, normalized integrated product

■ REFERENCES

(1) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: New York, 1989; pp 1−108.
(2) Leach, A. R. *Molecular Modelling: Principles and Applications*; Prentice Hall: Upper Saddle River, NJ, 2001; pp 165−406.
(3) Schlick, T.; Collepardo-Guevara, R.; Halvorsen, L. A.; Jung, S.; Xiao, X. Biomolecular modeling and simulation: A field coming of age. *Q. Rev. Biophys.* **2011**, *44*, 191−228.
(4) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120−127.
(5) Brooks, C. L. Protein and peptide folding explored with molecular simulations. *Acc. Chem. Res.* **2002**, *35*, 447−454.
(6) Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679−6685.
(7) Shea, J. E.; Brooks, C. L. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499−535.
(8) van Gunsteren, W. F.; Berendsen, H. J. C. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem., Int. Ed.* **1990**, *29*, 992−1023.
(9) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(10) Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. Challenges in protein-folding simulations. *Nat. Phys.* **2010**, *6*, 751−758.
(11) Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **2012**, *22*, 52−61.
(12) Zuckerman, D. M. Equilibrium sampling in biomolecular simulations. In *Annu. Rev. Biophys.*; Rees, D. C., Dill, K. A., Williamson, J. R., Eds; **2011**; Vol. *40*, pp 41−62.
(13) MacKerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.
(14) Villa, A.; Fan, H.; Wassenaar, T.; Mark, A. E. How sensitive are nanosecond molecular dynamics simulations of proteins to changes in the force field? *J. Phys. Chem. B* **2007**, *111*, 6015−6025.
(15) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, 3, 198−210.
(16) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(17) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(18) Grossfield, A.; Zuckerman, D. M. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* **2009**, *5*, 23−48.
(19) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To milliseconds and beyond: Challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58−65.
(20) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.
(21) Day, R.; Paschek, D.; Garcia, A. E. Microsecond simulations of the folding/unfolding thermodynamics of the trp-cage miniprotein. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1889−1899.
(22) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(23) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(24) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J. Biophys. Lett.* **2011**, *40*, 843−856.
(25) Reif, M. M.; Huenenberger, P. H.; Oostenbrink, C. New interaction parameters for charged amino acid side chains in the GROMOS force field. *J. Chem. Theory Comput.* **2012**, *8*, 3705−3723.
(26) MacKerell, A. D.; Feig, M.; Brooks, C. L. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(27) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(28) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(29) Li, D.-W.; Brueschweiler, R. NMR-based protein potentials. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778−6780.
(30) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(31) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(32) Sorin, E. J.; Pande, V. S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **2005**, *88*, 2472−2493.
(33) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. Force field bias in protein folding simulations. *Biophys. J.* **2009**, *96*, 3772−3780.
(34) Best, R. B.; Buchete, N.-V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **2008**, *95*, L7−L9.
(35) Matthes, D.; de Groot, B. L. Secondary structure propensities in peptide folding simulations: A systematic comparison of molecular mechanics interaction schemes. *Biophys. J.* **2009**, *97*, 599−608.
(36) Thompson, E. J.; DePaul, A. J.; Patel, S. S.; Sorin, E. J. Evaluating molecular mechanical potentials for helical peptides and proteins. *PLoS One* **2010**, *5*, e10056.
(37) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **2011**, *100*, L47−L49.
(38) Nerenberg, P. S.; Head-Gordon, T. Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides. *J. Chem. Theory Comput.* **2011**, *7*, 1220−1230.
(39) Beauchamp, K. A.; McGibbon, R.; Lin, Y.-S.; Pande, V. S. Simple few-state models reveal hidden complexity in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17807−17813.
(40) Kuhrova, P.; De Simone, A.; Otyepka, M.; Best, R. B. Force-field dependence of chignolin folding and misfolding: comparison with experiment and redesign. *Biophys. J.* **2012**, *102*, 1897−1906.
(41) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. Microsecond time-scale conformational exchange in proteins: Using long molecular dynamics trajectory to simulate nmr relaxation dispersion data. *J. Am. Chem. Soc.* **2012**, *134*, 2555−2562.
(42) De Sancho, D.; Best, R. B. What is the time scale for alpha-helix nucleation? *J. Am. Chem. Soc.* **2011**, *133*, 6809−6816.
(43) Lin, M. M.; Mohammed, O. F.; Jas, G. S.; Zewail, A. H. Speed limit of protein folding evidenced in secondary structure dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16622−16627.
(44) Saladino, G.; Marenchino, M.; Gervasio, F. L. Bridging the gap between folding simulations and experiments: The case of the villin headpiece. *J. Chem. Theory Comput.* **2011**, *7*, 2675−2680.
(45) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114*, 14916−14923.
(46) Best, R. B.; Mittal, J. Balance between alpha and beta structures in ab initio protein folding. *J. Phys. Chem. B* **2010**, *114*, 8790−8798.
(47) Wang, D.; Freitag, F.; Gattin, Z.; Haberkern, H.; Jaun, B.; Siwko, M.; Vyas, R.; van Gunsteren, W. F.; Dolenc, J. Validation of the GROMOS 54A7 force field regarding mixed alpha/beta-peptide molecules. *Helv. Chim. Acta* **2012**, *95*, 2562−2577.
(48) Shepherd, N. E.; Hoang, H. N.; Abbenante, G.; Fairlie, D. P. Single turn peptide alpha helices with exceptional stability in water. *J. Am. Chem. Soc.* **2005**, *127*, 2974−2983.
(49) Serrano, A. L.; Tucker, M. J.; Gai, F. Direct assessment of the alpha-helix nucleation time. *J. Phys. Chem. B* **2011**, *115*, 7472−7478.
(50) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing: Dordrecht, Holland, 1981; pp 331−342.
(51) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(52) Barker, J. A.; Watts, R. O. Monte Carlo studies of dielectric properties of water-like models. *Mol. Phys.* **1973**, *26*, 789−792.
(53) Smith, P. E.; van Gunsteren, W. F. Consistent dielectric properties of the simple point-charge and extended simple point-charge water models at 277 and 300 K. *J. Chem. Phys.* **1994**, *100*, 3169−3174.
(54) Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.
(55) Miyamoto, S.; Kollman, P. A. SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(56) Filipe, L. C. S.; Machuqueiro, M.; Baptista, A. M. Unfolding the conformational behavior of peptide dendrimers: Insights from molecular dynamics simulations. *J. Am. Chem. Soc.* **2011**, *133*, 5042−5052.
(57) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(58) Patriksson, A.; van der Spoel, D. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073−2077.
(59) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826−843.
(60) Piana, S.; Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(61) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(62) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer: New York, 2002; pp 1−9 and 111−133.
(63) Campos, S. R. R.; Baptista, A. M. Conformational analysis in a multidimensional energy landscape: Study of an arginylglutamate repeat. *J. Phys. Chem. B* **2009**, *113*, 15989−16001.
(64) Silverman, B. W. *Density Estimation for Statistics and Data Analysis*, 1st ed.; Chapman and Hall/CRC: New York, 1986; pp 7−94.

(65) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schuette, C.; Noe, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.
(66) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.
(67) Mu, Y. G.; Nguyen, P. H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 45−52.
(68) Buchete, N.-V.; Hummer, G. Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.
(69) Noe, F.; Schuette, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011−19016.
(70) Bowman, G.; Pande, V. Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10890−10895.
(71) Swope, W. C.; Pitera, J. W.; Suits, F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B* **2004**, *108*, 6571−6581.
(72) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput. Biol.* **2009**, *5*, e1000452.