High Temperature Unfolding Simulations of the TRPZ1 Peptide

Giovanni Settanni and Alan R. Fersht
Centre for Protein Engineering, Cambridge, United Kingdom

ABSTRACT We report high temperature molecular dynamics simulations of the unfolding of the TRPZ1 peptide using an explicit model for the solvent. The system has been simulated for a total of 6 μs with 100-ns minimal continuous stretches of trajectory. The populated states along the simulations are identified by monitoring multiple observables, probing both the structure and the flexibility of the conformations. Several unfolding and refolding transition pathways are sampled and analyzed. The unfolding process of the peptide occurs in two steps because of the accumulation of a metastable on-pathway intermediate state stabilized by two native backbone hydrogen bonds assisted by nonnative hydrophobic interactions between the tryptophan side chains. Analysis of the un/folding kinetics and classical commitment probability calculations on the conformations extracted from the transition pathways show that the rate-limiting step for unfolding is the disruption of the ordered native hydrophobic packing (Trp-zip motif) leading from the native to the intermediate state. But, the speed of the folding process is mainly determined by the transition from the completely unfolded state to the intermediate and specifically by the closure of the hairpin loop driven by formation of two native backbone hydrogen bonds and hydrophobic contacts between tryptophan residues. The temperature dependence of the unfolding time provides an estimate of the unfolding activation enthalpy that is in agreement with experiments. The unfolding time extrapolated to room temperature is in agreement with the experimental data as well, thus providing a further validation to the analysis reported here.

INTRODUCTION

The Trp-zip peptides (1) represent one of the smallest protein sequences that exhibit the characteristic folding features of a globular protein. Their size and microsecond folding kinetics (2) make them a system amenable to classical molecular dynamics simulations in full atomic detail (2–4). For the very same reasons, i.e., small size and fast kinetics, a thorough experimental characterization of the folding of the peptide is difficult. Three Φ-values are known for the TRPZ4 variant (5). These and other experimental data on the folding characteristics of these and related peptides, nicely synthesized in Du et al. (6), have demonstrated how the hydrophobic core composition affects mainly the unfolding rate of these hairpins, while the folding rate is mostly determined by the turn formation propensity of their sequence. These observations establish important features of the structure of the transition state of the peptides. However, the atomic details of this ensemble and the pathways leading to and from it as determined by a rigorous kinetic approach involving the use of molecular dynamics (MD) simulations with explicit treatment of the solvent are not yet available. For large proteins, when experimental data on the folding transition-state ensemble (TSE) (or any other state of interest) are available, usually consisting of an extensive Φ-value analysis, these can be used as constraints in the simulation to enrich the sampling of the phase space region of interest (7–10). In the case of TRPZ1, determination of the TSE is possible without the use of constraints because of the small size of the peptide.

Previous theoretical studies on the folding of Trp-zip peptides provide contrasting views. On the one hand, implicit water MD simulations of TRPZ2 using the replica exchange method (3) showed the presence of a corrugated free energy landscape with a single absolute minimum and many competing local minima at low and room temperature. On the other hand, distributed computing simulations using a different force-field and implicit treatment of the solvent (2) provided evidence for the presence of two distinct free energy minima corresponding to the folded and unfolded states. While simulations using implicit treatment of the solvent give access to the timescales needed to sample complete unfolding of the peptides, they neglect the active role that discrete water molecules may play in the folding process (11,12). This can only be captured by explicit treatment of the water. Zhang et al. performed replica exchange MD (REMD) calculations on TRPZ2 with explicit treatment of the water (4), which allowed them to elucidate the projected free energy landscape of the peptide as a function of several observables. Because REMD does not provide the correct kinetics of the simulated process, the authors partly based the search for the TSE of folding on an analysis of the saddle points in the projected free energy landscape. The latter procedure was shown to lead possibly to a wrong characterization of the activated species (12,13).

Owing to the very high activation free energy of unfolding of these peptides, it is unfeasible to sample their phase space

Submitted September 21, 2007, and accepted for publication January 3, 2008.

Address reprint requests to Giovanni Settanni, Tel.: +44-1223-402133; E-mail: gs@mrc-lmb.cam.ac.uk.

Editor: Angel E. Garcia.

0006-3495/08/06/4444/10 $2.00

doi: 10.1529/biophysj.107.122606

using straight room-temperature MD simulations with explicit treatment of the solvent. Several methods to improve the sampling of MD simulations have been proposed. These comprise sophisticated simulation technique like REMD (14,15), umbrella sampling (16,17), transition path sampling (18), the Markovian state model (19), and the metadynamics method (20,21).

High temperature unfolding simulations pioneered by Daggett and co-workers (22) have been extensively used for the dissection of the folding pathway and TSE of several proteins and benchmarked by comparison with experimental data on the TSE (23–29). Notwithstanding possible complications from limited sampling issues and the unphysical temperatures and pressures (or density) sampled along those simulations, this approach proved to be very successful in correctly identifying the structures of the TSE as compared to experiments, the exceptional agreement being probably related to the robust character of the free energy landscape of proteins, evolved or designed to exhibit a unique native structure (30).

In the following, in analogy to a temperature jump experiment, MD simulations of the TRPZ1 peptide with explicit treatment of the solvent have been performed raising the temperatures of the system at values above 373 K. By using standard observables and defining an instantaneous measure of the flexibility of the peptide (the sampled volume increase rate, VIR), we observed and analyzed several unfolding pathways and we made commitment probability calculations to map precisely peptide conformations to the free energy landscape of the peptide. The article addresses the following open questions:

- Do high temperature MD simulations of small peptides allow the reliable extraction of information on their TSE and folding mechanism?
- How well can experimental data be extrapolated from the simulations?
- Does the TRPZ1 peptide fold cooperatively?
- What is the atomic structure of the TSE of TRPZ1 peptide?
- How do the experimentally known determinants of the folding and unfolding rates of hairpin peptides affect the conformations of the populated states?

METHODS

Explicit water simulations

Simulations of the TRPZ1 peptide were performed using the GROMACS package (31) with the OPLSAA force field (32) and the SPC model (33) for the explicit treatment of the water molecules. The peptide in the NMR conformation (PDB ID: 1LE0) (1) was immersed in a periodic cubic box of 1666 water molecules. The minimal distance of the peptide from the initial box boundary is 0.8 nm. The peptide has been C-terminated with an NH2 group as in the NMR structure and two Cl^- ions have been added to neutralize the total peptide charge. The cutoff for the electrostatic and van der Waals interactions was set to 1 nm and the PME was used to take into account the long-range interactions. Before performing MD simulations, the system was minimized using steepest descent for 1000 steps. All the MD simulations were performed in the NPT ensemble with single precision and 2-fs integration time step. The LINCS algorithm (34) was used to restrain the length of all covalent bonds. The temperature and pressure have been regulated by the Berendsen algorithm (35) with coupling constants $\tau_t = 0.1$ ps and $\tau_p = 0.5$ ps, respectively. Several target values for the temperature have been used while the pressure has been kept constant at 1 atm in all simulations. Before the production phase, the water molecules were allowed to equilibrate for 10 ps at 300 K keeping the peptide fixed. Then, 10 runs were started of the whole system differing by the starting velocities initialized at 300 K. The runs were allowed to relax to the same target temperature along a 5-ns-long simulation. Target temperatures of 373 K, 400 K, 435 K, 450 K, and 470 K were used. Finally, the 10 runs were elongated for 100 ns each. The 373 K simulations where further elongated for a total of 200 ns per run. A 100-ns control run at 300 K was also performed. In the simulations done above the expected boiling temperature of water, during the equilibration phase, the system's density relaxes close to the vapor-liquid coexistence line for the SPC water model (see Fig. S10 in Supplementary Material, Data S1) as determined using Gibbs ensemble Monte Carlo simulations (36). The limited size of the system and the simulation protocol prevent the observation of the vapor-liquid phase transition (37). The system remains in the liquid state with no drift of the density along the simulations (see Fig. S11 in Data Sl). Table 1 summarizes the unfolding simulation setup. Snapshots of the system from the trajectories were saved every 20 ps for further analysis.

Cluster analysis

A cluster analysis was done on the concatenated trajectories at each temperature. The leader algorithm was used and the clustering was based on the $C_\alpha C_\beta$-distance root mean-square deviation (dRMSD) of the peptide (38). The cutoff for the clustering was set to 0.08 nm. The analyses that follow have also been done using 0.05 nm cutoff for the simulation at 450 K. The differences introduced by a smaller cutoff will be briefly analyzed. Cluster calculations were performed using the program WORDOM (39).

Explored conformational volume increase rate (VIR)

The leader algorithm for clustering allows for the partitioning of the conformational space of the peptide in a progressive way. New clusters, i.e., new partitions of the conformational space, are defined when the trajectory leaves the conformational space that has already been sampled and partitioned. Each cluster represents a portion of a sphere in the multidimensional space of the peptide where the radius is the cutoff distance used for clustering. When the trajectory is sampling conformations from the native state of a protein or peptide, characterized by a well-defined conformational energy minimum and low entropy, it will enter a limited number of conformational space partitions (i.e., clusters). Thus, after a lag time due to diffusion, no new clusters will be visited and the explored VIR will be close to zero. On the other hand, as the system crosses the folding free energy barrier and reaches the denatured state (assumed to be a state characterized by a large entropy), the explored volume (i.e., the number of new clusters visited by the trajec-

TABLE 1 Unfolding simulation setup

Temperature (K)	No. simulations	Length (ns)
373	10	200
400	10	100
435	10	100
450	10	100
470	10	100

tory) will start to increase linearly with the simulation time, as the peptide diffuses in the new state (Fig. S12 in Data S1). Given the high dimensionality of the conformational space and the large entropy of the denatured state, saturation could only be achieved by extremely long MD simulations far from the reach of current computer power. However, the progressive partitioning of the conformational space along an unfolding simulation allow the measurement of the VIR and, consequently, picking the unfolding event quite precisely. A similar quantity has been used for the assessment of the speed of convergence of different simulation methods (21). The raw increase in number of sampled clusters was filtered by convolution with a Gaussian function with 0.1-ns standard deviation to eliminate fast fluctuations. The VIR is the derivative of this measure. The same filter was applied to root mean-square deviation (RMSD) and Q data. A filter with 0.4-ns width was also tested, providing small changes to the outcomes of the analysis.

Putative transitions

The distribution of the VIR, number of native backbone hydrogen bonds Q, and C_αRMSD from the native state of the conformations sampled during the unfolding simulations presents three populated regions at most of the temperatures that have been sampled. Derived from unfolding simulations, these distributions are not the equilibrium distributions; however, they indicate the populated states of the peptide (Fig. 1). The two conformationally restricted states (i.e., VIR $< 100\ ns^{-1}$) are the only populated regions in the simulations done at 373 K. They are both characterized by the presence of different amounts of native structure (low RMSD, $Q > 0$). The native state (N) has $Q > 3.2$ and RMSD < 0.15 nm, the intermediate state (I) has $Q \approx 2$ ($1 < Q < 2.2$) and RMSD ≈ 0.2 nm (0.1 nm $<$ RMSD $<$ 0.35 nm). The high VIR peak conformations (i.e., VIR $> 300\ ns^{-1}$) are associated with the denatured state (D), and present non-native-like characteristics (C_αRMSD > 0.3 nm from the native state and $Q < 0.5$). Average transition times from one state to the other were collected as follows. The total dwelling time in each state was divided by the number of observed transitions into each other state. This procedure takes into account also the stretches of trajectories with no commitment and provides a ''maximum-likelihood'' estimate of the transition time assuming first-order transitions (40,41). This assumption was tested by examining the distribution of transition times where the number of observed transition was large enough to allow for it (>10). In these cases, the distribution was found to be exponential, as expected, with level of confidence >0.01, according to χ^2 tests. A reduction of the clustering radius from 0.08 nm to 0.05 nm, results in a shift of the I-state population to a region with slightly larger VIR values, while the D state is flattened to the maximal VIR value, which depends on the frequency for saving the trajectory coordinates. The change, however, has no influence on the results of the following analysis.

Validation of transition states by P_{commit} analysis

From the productive trajectories, i.e., the stretches of trajectory connecting the most populated states (N,I,D), several snapshots were selected and P_{commit} was computed for each. The selection process involved both blind selection of evenly-spaced conformations along the transition pathway, and, in some cases, a bisection procedure aimed at identifying the transition state along the pathway. P_{commit}^{N-I} and P_{commit}^{I-D} were computed according to the pair of states connected by each productive trajectory. The P_{commit} calculations consisted of 20 MD runs from the initial conformation with varying starting velocity. Each run was 2-ns long. The P_{commit} is the fraction of runs committed to one state (conventionally the most nativelike). Several strategies could be adopted to classify the runs as committed to a state. Here, to avoid uncommitted runs, the conformational space of the peptide has been divided into two regions by a separatrix passing by the saddle point between the two states in the VIR-RMSD-Q projection of the unfolding simulations (42). Along each run, the simulation can cross the separatrix several times. The fraction of runs on one side of the separatrix is a function of the time elapsed from the beginning of the runs, $P(t)$. Generally, $P(t)$ relaxes to a certain value within the time length of the runs (i.e., 2 ns, Fig S15 in Data S1). Here, the assumption is made that the relaxation occurs exponentially and $P(t)$ is fitted with an exponential function. P_{commit} is defined as the asymptotic value of the

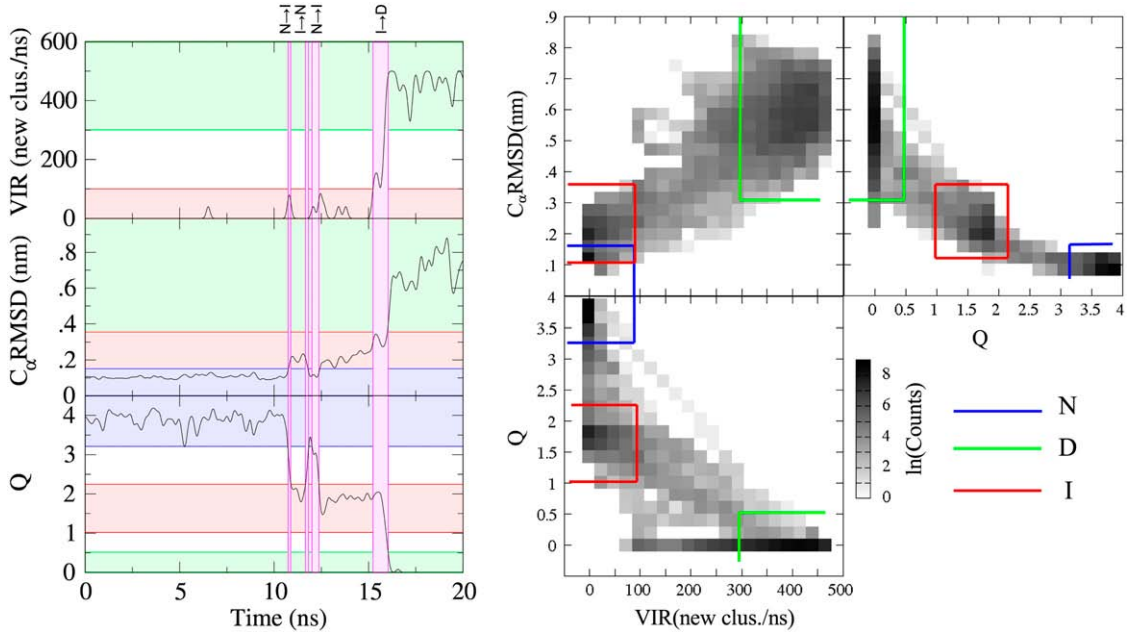

FIGURE 1 Identification of transition pathways. On the left, sample time series of VIR, C_α-RMSD, and Q are reported. On the right, the projection of the time series on two-dimensional density plots allow us to identify the most populated regions of the parameter space. Green, red, and blue lines mark the borders of the D, I, and N states, respectively. The boundary values of the observables defining the states are then reported on the time series on the left. The stretches of trajectory where all the observables at the same time cross from one state to another, are marked as transition pathways (shaded in *magenta*).

exponential fit. Other definitions of P_{commit} are possible and have been applied to the present data (not shown). Generally they provide similar results to those obtained with the above method; however, the latter is more robust to slight changes in the definition of the states, because it allows us to average-out the time-dependent behavior of $P(t)$.

RESULTS

Populated states during unfolding

The NMR conformation (1) represents a free energy minimum for the force field used in this work both in the control run at room temperature and in the unfolding simulations at high temperature. As anticipated in Methods, three density peaks were observed in the projection of the trajectories on the VIR-Q-RMSD space (Fig. 1, *right*). This picture is confirmed by a dRMSD-based cluster analysis of the trajectories (Fig. 2). The dwelling times in each state has been reported in Table 2. As expected, the larger the temperature the smaller the amount of time the peptide spends in the N state, both because of the decreasing stability of N and because of faster kinetics. Interestingly, apart from the results obtained at 373 K (where D is not reached), the population of I is always smaller than D. A detailed analysis of the I state shows that the residual structure consists of the backbone hydrogen bonds between the carboxy oxygen of Thr^3 and the amide hydrogen of Thr^{10} and between the amide hydrogen of Glu^5 and carboxy oxygen of Lys^8. Further, the ordered Trp-zip motif is lost in the I state. The distances between the Trp residues on the same hairpin strand decrease, while those between Trp residues on opposite strands increase, with the exception of the pair Trp^4 and Trp^9, indicating the formation of an extended hydrophobic cluster that is less ordered than the zip motif (see Fig. S13 in Data S1). Several populated rotamers for the Trp side chains are present in the I and N states (see Fig. S13 in Data S1). However, the exchange rate between these rotamers is considerably faster than the rate of exchange among the states I, N, and D (see Fig. S14 in Data S1).

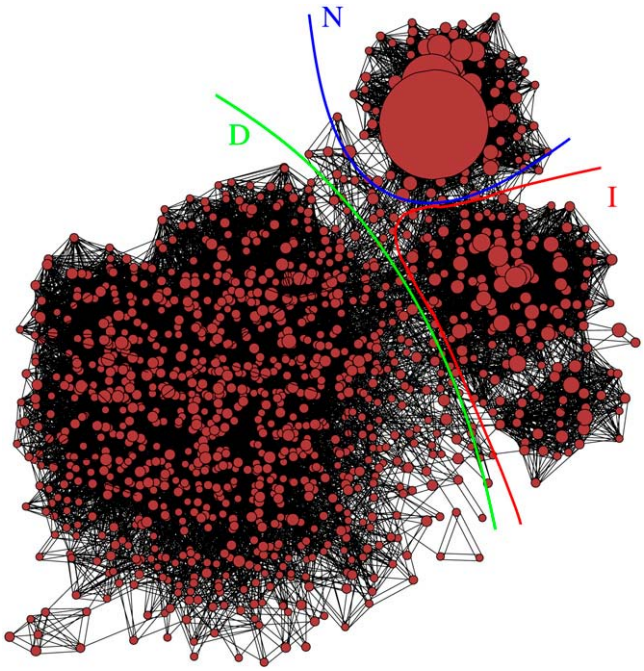

FIGURE 2 Kinetic network representation of the sampled conformational space of TRPZ1 at 450 K. Each circle represents a cluster obtained using the leader algorithm based on a DRMS metric between conformations. The size of the circle is proportional to the number of structures in the cluster. Pair of clusters are connected by lines when a transition has been sampled along the simulations. Clusters with similar connectivity pattern are placed close together in the plot. The large cluster on the top of the figure and the connected ancillary clusters within the blue boundary corresponds to the native state N. The large set of small clusters evenly distributed on the left of the figure within the green boundary corresponds to the D state. The couple of sets of clusters below the Native state and within the red boundary corresponds to the I state. The I state provides the more frequent connection between the N and D states, as revealed by the density of connecting lines. For graphical reason, small clusters were lumped together according to the dRMSD between cluster centers, so that, at the end of the procedure, all the clusters contain more than 20 conformations. Please note that the VIR was computed using unlumped clusters. The figure was prepared using VISONE (http://visone.info), similarly to what was done in Rao and Caflisch (48).

Transition events

Few direct transitions were observed between N and D that had no significant dwelling times spent in I (25% of the total number of transitions with no change of this fraction with temperature), indicating that I is a predominant on-pathway intermediate (Fig. 3). The magnitudes of the transition times between the states as a function of temperature (Table 3) provide much information. At 373 K, only few transitions to I and back to N were observed and the dwelling time in the I state was very small, indicating a poor sampling of this state. At the higher temperatures, transitions to D were also observed. In those cases, the slowest unfolding process is the N→I transition. The completion of the unfolding process by the I→D transition occurs on a shorter timescale. As expected in high temperature unfolding simulations, statistics for the folding process is scarcer than for the unfolding process. In the case of temperatures above 435 K, the statistics is sufficient to draw a qualitative picture of the underlying free energy landscape of the peptide (Fig. 4). As expected in nonequilibrium unfolding simulations, the measured population of each state (Table 2) does not match information on the relative free energy as measured by kinetic rates (Fig. 4). The N state is overpopulated because all the simulations were started from that state. However, the I state is correctly found with a population smaller than D.

TABLE 2 Populations of states

Temperature	P(N)	P(I)	P(D)
373 K	0.983	0.006	0.000
400 K	0.806	0.039	0.111
435 K	0.358	0.138	0.472
450 K	0.271	0.140	0.545
470 K	0.139	0.075	0.769

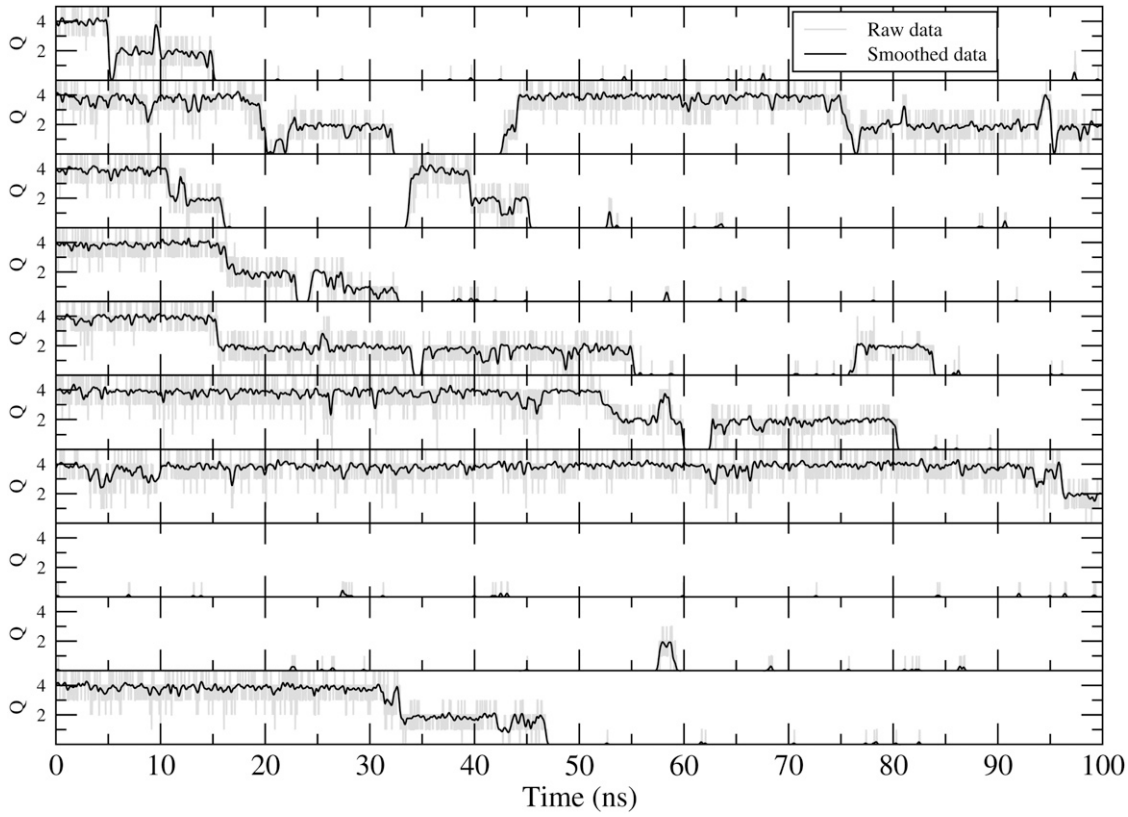

FIGURE 3 Time series of Q along the 10 trajectories run at 450 K. The system spends considerable amounts of time in the native state N ($Q > 3.2$) and in the intermediate state I ($1 < Q < 2.2$) before reaching the D state ($Q < 0.5$). Transitions between states occur sharply, and not as slow diffusional processes. Only few occurrences of direct transitions from N to D can be observed (i.e., transitions between N and D with no significant dwelling time in the I state).

The kinetics of unfolding was sampled sufficiently to construct an Eyring plot for the reaction (Fig. 5), which shows that unfolding follows the Arrhenius law, as expected (43). The slope of the regression line gives an activation enthalpy for unfolding as $\Delta H^{\ddagger} = 14.0 \pm 1.2$ Kcal/mol. The extrapolation of the unfolding time to room temperature gives an estimate for the unfolding rate constant $k_u = (62\ \mu s)^{-1}$.

TABLE 3 Transition times between states

T (K)	$n_{N\to I}$	$\langle t_{N\to I}\rangle$ (ns)	$n_{I\to N}$	$\langle t_{I\to N}\rangle$ (ns)
373	4	450.56 ± 225.28	4	2.93 ± 1.46
400	4	210.71 ± 105.36	1	42.24 ± 42.24
435	15	24.68 ± 6.37	7	21.49 ± 8.12
450	18	16.16 ± 3.81	8	19.37 ± 6.85
470	16	8.74 ± 2.19	7	11.23 ± 4.24
	$n_{I\to D}$	$\langle t_{I\to D}\rangle$ (ns)	$n_{D\to I}$	$\langle t_{D\to I}\rangle$ (ns)
373	0	—	0	—
400	2	21.12 ± 14.93	0	111.26*
435	16	9.40 ± 2.35	9	53.15 ± 17.72
450	20	7.75 ± 1.73	12	46.11 ± 13.31
470	16	4.91 ± 1.23	7	106.05 ± 40.08

*This number is the total dwelling time in the D state observed at that temperature. It represents a lower bound estimate for the transition time since no transition of this kind was observed.

Limited sampling, especially at 373 K and 400 K, prevents the same analysis for the folding process. Estimate of the population of I with respect to all the other states needs a precise knowledge of all the transition rates; thus, we cannot reliably measure it at the lower temperatures or, worse, extrapolate it to room temperature.

Identification and characterization of the transition states for folding

Given the relatively small number of conformations sampled along the productive trajectories connecting the stable states N, I, and D, it was possible to measure the P_{commit}^{N-I} and P_{commit}^{I-D} of many of them and so identify the transition states at $T =$ 400 K, 435 K, and 450 K. Because of the smaller number of transition pathways available at 400 K, only 40 conformations were selected for P_{commit} calculations, while 115 and 90 conformations were selected at 435 K and 450 K, respectively. Both P_{commit}^{N-I} and P_{commit}^{I-D} generally decrease along the unfolding transition pathways in a sigmoidal way and in the same way they increase along the folding transition pathways. At 450 K, large changes in P_{commit} occur, usually, on a relatively narrow time interval of ~0.2 ns for both the N→I and I→D transition, while the P_{commit} before and after the

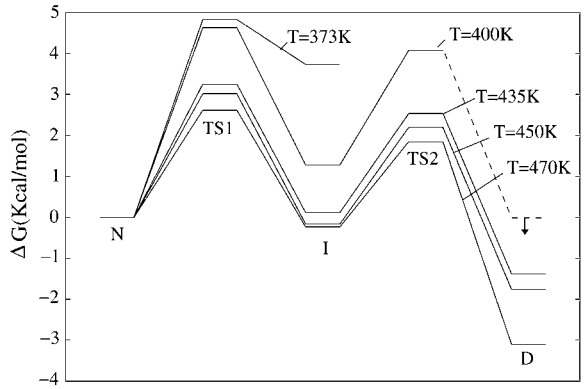

FIGURE 4 A qualitative picture of the free energy landscape of the TRPZ1 peptide determined on the basis of the transition times between states at different temperatures. TS1 and TS2 indicate the transition states for the N→I and I→D transitions, respectively. The ΔG associated with each barrier has been determined using the formula $k = a\exp(-\Delta G/RT)$ where R is the gas constant, T is the temperature, k is the measured rate for the transition, and $a = bT$ is the preexponential factor that has been set to $(0.5\ \text{ns})^{-1}$ at 450 K and scaled accordingly at the other temperatures. This value for the preexponential factor has been chosen on the basis of the time measured to observe a complete shift in P_{commit} (from 0 to 1 or from 1 to 0) along the trajectory for the I→D transition. Please note that, while the height of the barriers depends on the choice of the preexponential factor, the free energy difference between minima is independent of it. The dashed line indicates that no transition was observed in that case; thus, the corresponding rate is supposed to be smaller than the inverse dwelling time spent by the system in the unproductive state. Error estimates for the free energy differences are usually comprised between 0.2 and 0.5 Kcal/mol, as determined by propagation of the uncertainties on the transition times (Table 3), with the exception of the free energy of TS2 and D at 400 K where the propagated error is 0.9 and 1.0 Kcal/mol, respectively.

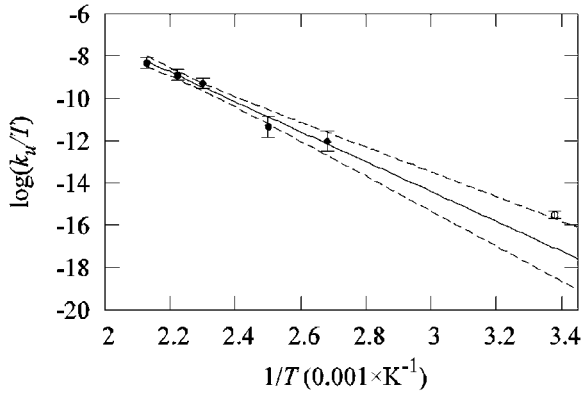

FIGURE 5 Eyring plot based on the data coming from the simulations (*solid circles*). The linear fit of the $\log(k_u/T)$ as a function of $1/T$ is shown (*solid line*). The slope provides an estimate for the unfolding activation enthalpy $\Delta H^{\ddagger} = 14.0 \pm 1.2$ Kcal/mol. The fit is extrapolated to room temperature. The 95% credibility interval for the extrapolation is provided (*dashed lines*). The available experimental data (*open circle*) correspond to an unfolding time of 18.3 ± 3.1 μs at $T = 296$ K (2). The unfolding time as extrapolated from the fit is 62 μs.

transition is almost constantly either 0 or 1, as expected. The time width of the transition is weakly dependent on temperature, thus at 400 K it is generally larger (≈0.3 ns) than at 450 K, while its variability at a given temperature is quite large since it can range from 0.06 ns to 0.8 ns. In some relatively long transition pathways, $P^{\text{N-I}}_{\text{commit}}$ is not monotonic, evidencing some diffusional phenomenon occurring on the top of the free energy barrier. As mentioned in Methods, both $P^{\text{N-I}}_{\text{commit}}(t)$ and $P^{\text{I-D}}_{\text{commit}}(t)$ usually converge to values close to the $P^{\text{N-I}}_{\text{commit}}$ and $P^{\text{I-D}}_{\text{commit}}$ values, respectively, with an exponential relaxation. The characteristic time of this process is usually ~0.2 ns at 450 K and increase to 0.4 ns at 400 K. Slower relaxations with up to 0.8 ns characteristic time were rarely observed.

N→I transition

The validated transition pathways for the N→I transition (Fig. 6) show that the Trp-zip motif is lost just before reaching the transition point. On the other hand, the conformation of the turn can be rather non-native-like in the TSE. The backbone hydrogen bond between Thr^3 and Thr^{10} is preserved along the pathway. The other native backbone hydrogen bonds (i.e., in distal or proximal position) may or may not be present, indicating that their formation is not rate-limiting. Trp^2 and Trp^9 are in a parallel configuration in most of the TS conformations (Fig. 6 *b*), while in conformations with larger $P^{\text{N-I}}_{\text{commit}}$, they assume the characteristic interlocked configuration with the other two Trp residues.

The degree of native order of the Trp-zip motif was monitored using the difference in the distance between the Cγ atoms of the pairs of residues Trp^4-Trp^2 and Trp^4-Trp^9 (D_{zip}). In the native state, D_{zip} is large and positive, while it decreases with $P^{\text{N-I}}_{\text{commit}}$ and is usually negative for conformation in the I state ($P^{\text{N-I}}_{\text{commit}} = 0$, Fig. 7). Q or RMSD from the NMR conformations are not as well correlated to $P^{\text{N-I}}_{\text{commit}}$, neither are other observables probing unordered hydrophobic collapse (e.g., the exposed hydrophobic surface area) or turn formation (e.g., hydrogen bonds proximal to the turn). Thus, the data presented here point to the disruption of the native hydrophobic packing as the main rate-limiting step of the unfolding reaction. TSE conformations at 450 K are more nativelike than those at 435 K, as reported by the D_{zip} in agreement with Hammond behavior. Few TSE conformations at 400 K have been identified, so this does not allow us to make a similar statement for such a temperature.

I→D transition

The validated transition pathways for the I→D transition show that, in this case, the loss of the backbone hydrogen bond between Thr^3 and Thr^{10} is the rate-limiting step (Figs. 8 and 9). This process is concurrent with the loss of nonnative unspecific hydrophobic interactions between the side chains of Trp residues on the opposite sides of the hairpin. Conversely, the degree of nativeness of the turn region (residues

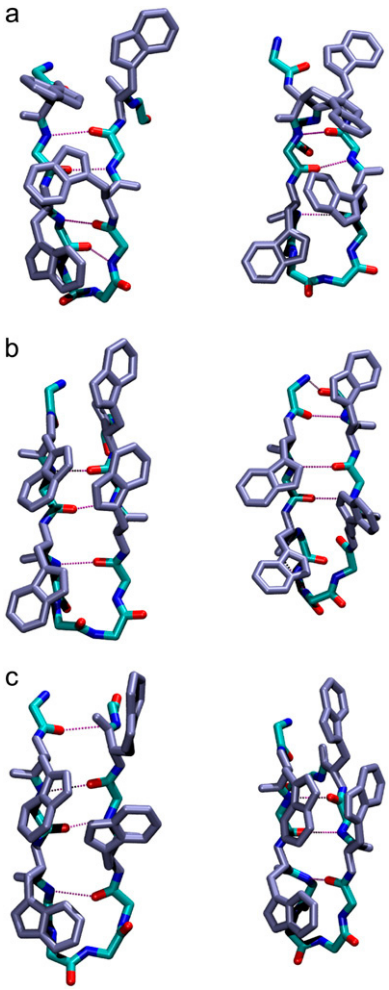

FIGURE 6 Conformations form several transition pathways of the simulations run at 450 K. (*a*) Conformations with $P^{N-I}_{commit} > 0.7$. (*b*) Conformations with $P^{N-I}_{commit} \approx 0.5$. (*c*) Conformations with $P^{N-I}_{commit} < 0.2$. The N-terminal strand is located on the left of each picture. The Trp-zip motif can be considered as natively ordered when Trp^9 on the C-terminal strand is closer to Trp^4 than Trp^2 on the N-terminal strand like it is obvious in panel *a*. At TSE (*b*) the native Trp-zip motif is lost. Conformations in panel *c*, that have low P^{N-I}_{commit}, already belong to the I state. In some cases (*c* on the *left*), water molecules disrupting the hydrophobic cluster formed by Trp^2, Trp^4, and Trp^9 contribute to the low P^{N-I}_{commit} notwithstanding the relative position of the Trp side chains.

5–7) as measured by its backbone RMSD is less correlated to the P^{I-D}_{commit} (data not shown). Taken together, these data indicate that loop closure is rate-determining in the D→I transition. Namely, the transition can occur either as a zipping process initiating in the turn region or by diffusional encounter of the termini (Fig. 8).

DISCUSSION

The unfolding simulations presented in this study directly address the kinetics of folding of the TRPZ1 peptide using an accurate and explicit treatment of the solvent degrees of

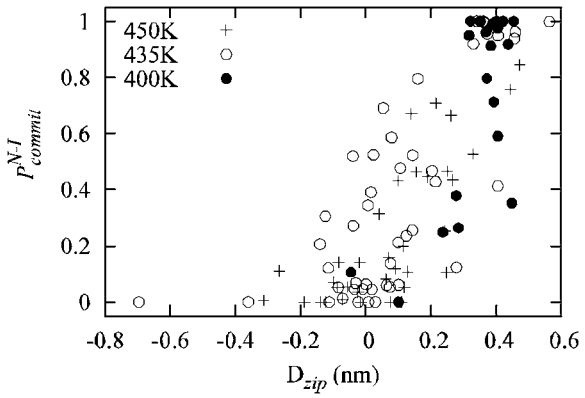

FIGURE 7 P^{N-I}_{commit} as a function of the degree of formation of the Trp-zip motif, as probed by the difference in distance between Cγ atom pairs of Trp^4-Trp^2 and Trp^4-Trp^9 (D_{zip}). Large values in this quantity are characteristic of the ordered native zip-motif. The motif order is progressively lost (small or negative values in the difference) as the free energy barrier for unfolding is crossed. At 450 K the conformations with a given P^{N-I}_{commit} have a more ordered zip-motif than at 435 K, indicating Hammond behavior. Data at 400 K are scarce to draw a conclusion. D_{zip} reports about the order of the zip motif only in the vicinity of the N state (i.e., when both *Q* and RMSD reports values close to nativelike). In regions clearly further from the N state D_{zip} is not supposed to provide the information. The most external points present in the low-P^{N-I}_{commit}-high-D_{zip} region correspond to conformations where water molecules are infiltrating the hydrophobic cluster formed by residues Trp^2, Trp^4, and Trp^9 (see Fig. 6 *c*).

freedom. Values of the activation enthalpy for unfolding ($\Delta H^{\ddagger} = 14.0 \pm 1.2$ Kcal/mol) and the extrapolation of the unfolding rate to room temperature ($k_u = (62\ \mu s)^{-1}$ are in good agreement with the available experimental data of 13.2 ± 0.5 Kcal/mol and $(18\ \mu s)^{-1}$ (2,3), respectively.

The simulations presented here provide evidence for the presence of a metastable intermediate state at any simulated temperature where unfolding occurs. A similar state has been previously observed in explicit water REMD simulations of TRPZ2 (state P in (4)). In that case, however, the analysis of the free energy landscape projected on various reaction variables does not indicate the N→P as the rate-determining step for unfolding. This could be either due to the fact that no trajectory smoothing was used on the number of backbone hydrogen bonds (which in the case of a discrete variable like *Q* eventually may lead to an artificial lumping of states with overlapping boundaries, as shown in Fig. S16 in Data S1), or to differences in the force fields. The metastability of I may also explain the large population of clusters of the TSE and the relatively large lifespan of the TSE as identified by Zhang et al. (4). Simulations performed on TRPZ1 and TRPZ2 using implicit treatment of the solvent, on the contrary, do not report such an intermediate state (2,3). However, a populated state with characteristics similar to the one observed in this work is apparent in the REMD work by Yang et al. (3). In that work, the emergence of free energy minima alternative to the native state, is interpreted as a signal for heterogeneous folding. The present simulations, instead, show a clear evidence that the folded state (N) remains a free energy mini-

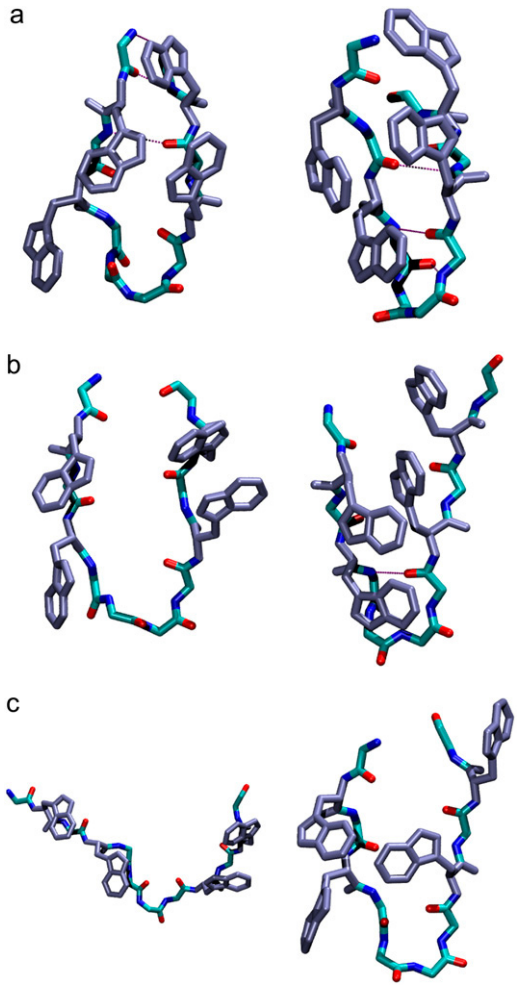

FIGURE 8 Conformations form two transition pathways of the simulations run at 450 K. (*a*) Conformations with $P^{D-I}_{commit} > 0.7$. (*b*) Conformations with $P^{D-I}_{commit} \approx 0.5$. (*c*) Conformations with $P^{D-I}_{commit} < 0.2$. The N-terminal strand is located on the left of each picture. Closure of the loop between Thr^3 and Thr^{10} and the Trp residues is rate-limiting in this case. This can occur either as a zipping process starting from the proximal part of the hairpin (*right column*) or as a process driven by diffusion of the termini, starting from the distal part of the hairpin (*left column*). Please note that, in the transition state conformation on the left, no intervening water molecule is present between the Trp^2 and Trp^{11} side chains and between the methyl groups of the Thr^3 and Thr^{10} side chains, which, for simplicity, were not shown. Thus, even in that case, the loop closure is taking place at the rate-limiting step.

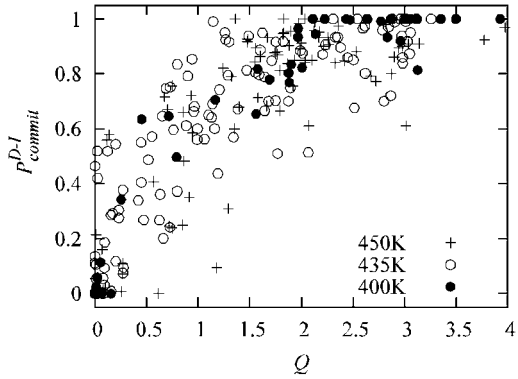

FIGURE 9 P^{I-D}_{commit} versus Q (after smoothing). In the transition between I and D, Q reports mainly the formation of the hydrogen bonds between Thr^3 and Thr^{10} and between Thr^5 and Lys^8. The correlation between Q and P^{I-D}_{commit} increases with decreasing temperature.

mum even at very high temperature and that the I state is only metastable. Simulations performed on a related peptide, the GB1 hairpin from the C-terminal domain of protein G, also implicated the presence of metastable intermediate states on the unfolding pathways (15,44,45). The I state found in this work resembles the F state of GB1 where interactions between the hydrophobic core residues and backbone hydrogen bonds close to them are present while the distal part of the hairpin and some of the proximal hydrogen bonds are lost. In the present simulations, however, the H state found in GB1, where only residual hydrophobic interaction are present with no backbone hydrogen bonds, exchanges too quickly with the completely unfolded state to be considered metastable. A further difference from GB1 is that the slowest process in TRPZ1 unfolding is the disruption of the native hydrophobic packing (N→I transition) while in GB1 the F→H transition is considered to be rate-determining (45). Currently there are no reports of experimental data describing the presence of the I state. The metastability of I and the small size of the spectroscopic signals associated with this peptide (C. M. Johnson, personal communication, 2007) may have precluded its experimental detection.

The MD simulations of TRPZ1 provide evidence that the rate-determining steps for the unfolding consist in the disruption of the native packing of the hydrophobic core (i.e., transition from N to I). Namely, in the sequential unfolding of the peptide from N to D via I, the N→I transition is the slowest step (Table 3). This is in agreement with the observations made by Du et al. (6), that TRPZ4 and GB1 peptides, two hairpins differing only in the hydrophobic core residues, have very different unfolding rates and similar folding rates, meaning that native hydrophobic core interactions are lost at the TSE, as has been shown in this work. In the folding process, instead, loop closure (i.e., the transition from D→I) mainly contributes to the determination of the rate. Namely, in the sequential refolding of the peptide from D to N via I, the D→I transition is the slowest step (Table 3). This process depends on several factors, mainly the turn formation propensity of the sequence (46) and the length of the loop to be closed. An analysis of the kinetics of several Trp-zip peptides differing only in the turn sequence (6) has shown that, while their unfolding rates are similar (within a factor of 3 from each other), their folding rates may change by a factor of 30 and their logarithm correlate very well with the entropic change upon folding. Since the entropy of the native state of peptides of similar length and sequence can reasonably be assumed to be small and not to change significantly across the peptides, turn formation propensity affects mainly the entropy of the unfolded state. Peptides with rather rigid turn sequences like the TRPZ3 (1) with ^{D}Pro residue (47) at the

turn will experience a much smaller accessible conformational space in the unfolded state than peptides with more flexible turns. A small entropy of the unfolded state translates into a small free energy penalty upon hairpin loop closure, and, eventually a small folding time. Namely the folding time of TRPZ3 is the smallest (1.7 μs) recorded among Trp-zip peptides, as noted by Du et al. (6).

CONCLUSIONS

The simulations presented in this work show that the unfolding of TRPZ1 at high temperature is a barrier-limited activated process. Agreement with available experimental data on the enthalpy of folding and unfolding rates confirm the validity of this analysis. Further they indicate the presence of a metastable on-pathway intermediate between the folded and the completely unfolded states. The rate-limiting step of the unfolding process is the transition between the native and the intermediate state, involving the disruption of the native hydrophobic packing (the ordered Trp-zip motif). The timescale of the folding process, instead, is mainly determined by transition from the completely unfolded state and the intermediate that involves the closure of the loop between Thr^3 and Thr^{10} concurrent to hydrophobic collapse of the two strands. This study provides a mechanistic picture that explains available experimental data on the folding of Trp-zip peptides and related β-hairpins.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

We thank Chris Johnson for critical reading of the manuscript.

G.S. has been partly funded by the Stiefel-Zangger-Stiftung fellowship, by the Swiss National Science Foundation fellowship for young investigators, and by the Novartis foundation.

REFERENCES

1. Cochran, A. G., N. J. Skelton, and M. A. Starovasnik. 2001. Tryptophan zippers: stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. USA*. 98:5578–5583.
2. Snow, C. D., L. Qiu, D. Du, F. Gai, S. J. Hagen, and V. S. Pande. 2004. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc. Natl. Acad. Sci. USA*. 101:4077–4082.
3. Yang, W. Y., J. W. Pitera, W. C. Swope, and M. Gruebele. 2004. Heterogeneous folding of theTrp-zip hairpin: full atom simulation and experiment. *J. Mol. Biol.* 336:241–251.
4. Zhang, J., M. Qin, and W. Wang. 2006. Folding mechanism of β-hairpins studied by replica exchange molecular simulations. *Proteins*. 62:672–685.
5. Du, D. G., M. J. Tucker, and F. Gai. 2006. Understanding the mechanism of β-hairpin folding via ϕ-value analysis. *Biochemistry*. 45:2668–2678.
6. Du, D., Y. Zhu, C.-Y. Huang, and F. Gai. 2004. Understanding the key factors that control the rate of β-hairpin folding. *Proc. Natl. Acad. Sci. USA*. 101:15915–15920.
7. Vendruscolo, M., E. Paci, C. M. Dobson, and M. Karplus. 2001. Three key residues form a critical contact network in a protein folding transition state. *Nature*. 409:641–645.
8. Paci, E., M. Vendruscolo, C. M. Dobson, and M. Karplus. 2002. Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* 324:151–163.
9. Paci, E., J. Clarke, A. Steward, M. Vendruscolo, and M. Karplus. 2003. Self-consistent determination of the transition state for protein folding: application to a fibronectin type III domain. *Proc. Natl. Acad. Sci. USA*. 100:394–399.
10. Settanni, G., J. Gsponer, and A. Caflisch. 2004. Formation of the folding nucleus of an SH3 domain investigated by loosely coupled molecular dynamics simulations. *Biophys. J.* 86:1691–1701.
11. Cheung, M. S., A. E. Garcia, and J. N. Onuchic. 2002. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA*. 99:685–690.
12. Bolhuis, P. G. 2005. Kinetic pathways of β-hairpin (un)folding in explicit solvent. *Biophys. J.* 88:50–61.
13. Rao, F., G. Settanni, and A. Caflisch. 2007. Estimation of folding probabilities and ϕ-values from molecular dynamics simulations of reversible peptide folding. *Methods Mol. Biol.* 350:225–249.
14. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.
15. Garcia, A. E., and K. Y. Sanbonmatsu. 2001. Exploring the energy landscape of a β-hairpin in explicit solvent. *Proteins*. 42:345–354.
16. Brooks, C., and D. A. Case. 1993. Simulations of peptide conformational dynamics and thermodynamics. *Chem. Rev.* 93:2487–2502.
17. Boczko, E. M., and C. L. I. Brooks. 1995. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science*. 269:393–396.
18. Bolhuis, P. G., D. Chandler, C. Dellago, and P. L. Geissler. 2002. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53:291–318.
19. Singhal, N., C. D. Snow, and V. S. Pande. 2004. Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper β-hairpin. *J. Chem. Phys.* 121: 415–425.
20. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*. 99:12562–12566.
21. Bussi, G., F. L. Gervasio, A. Laio, and M. Parrinello. 2006. Free-energy landscape for β-hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* 128:13435–13441.
22. Li, A., and V. Daggett. 1994. Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA*. 91:10430–10434.
23. Daggett, V., A. J. Li, L. S. Itzhaki, D. E. Otzen, and A. R. Fersht. 1996. Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* 257:430–440.
24. Mayor, U., C. M. Johnson, V. Daggett, and A. R. Fersht. 2000. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. USA*. 97:13518–13522.
25. Kazmirski, S. L., K. B. Wong, S. M. V. Freund, Y. J. Tan, A. R. Fersht, and V. Daggett. 2001. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci. USA*. 98:4349–4354.
26. Fersht, A. R., and V. Daggett. 2002. Protein folding and unfolding at atomic resolution. *Cell*. 108:573–582.
27. Mayor, U., N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. V. Freund, D. O. V. Alonso, V. Daggett, and A. R. Fersht. 2003. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*. 421:863–867.
28. Petrovich, M., A. L. Jonsson, N. Ferguson, V. Daggett, and A. R. Fersht. 2006. ϕ-Analysis at the experimental limits: mechanism of β-hairpin formation. *J. Mol. Biol.* 360:865–881.

29. Interlandi, G., G. Settanni, and A. Caflisch. 2006. Unfolding transition state and intermediates of the tumor suppressor p16(ink4a) investigated by molecular dynamics simulations. *Proteins.* 64:178–192.

30. Vendruscolo, M., E. Paci, M. Karplus, and C. M. Dobson. 2003. Structures and relative free energies of partially folded states of proteins. *Proc. Natl. Acad. Sci. USA.* 100:14817–14821.

31. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:306–317.

32. Jorgensen, W., D. Maxwell, and J. Tirado-Rives. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118: 11225–11236.

33. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Intermolecular Forces. B. Pullman, editor. Reidel, Dordrecht, the Netherlands.

34. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

35. Berendsen, H., J. Postma, W. van Gunsteren, A. DiNola, and J. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

36. Errington, J., and A. Panagiotopoulos. 1998. A fixed point charge model for water optimized to the vapor-liquid coexistence properties. *J. Phys. Chem. B.* 102:7470–7475.

37. Frenkel, D., and B. Smit. 2001. Understanding Molecular Simulation. Academic Press, Orlando, FL.

38. Settanni, G., F. Rao, and A. Caflisch. 2005. Φ-Value analysis by molecular dynamics simulations of reversible folding. *Proc. Natl. Acad. Sci. USA.* 102:628–633.

39. Seeber, M., M. Cecchini, F. Rao, G. Settanni, and A. Caflisch. 2007. WORDOM: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics.* 23:2625–2627.

40. Zagrovic, B., C. D. Snow, M. R. Shirts, and V. S. Pande. 2002. Simulation of folding of a small α-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* 323: 927–937.

41. Jayachandran, G., V. Vishal, and V. S. Pande. 2006. Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece. *J. Chem. Phys.* 124:1094–1107.

42. Snow, C. D., Y. M. Rhee, and V. S. Pande. 2006. Kinetic definition of protein folding transition state ensembles and reaction coordinates. *Biophys. J.* 91:14–24.

43. Fersht, A. R. 1999. Structure and mechanism. *In* Protein Science: Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman, New York, NY.

44. Zhou, R., B. J. Berne, and R. Germain. 2001. The free energy landscape for β-hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA.* 98:14931–14936.

45. Bolhuis, P. G. 2003. Transition-path sampling of β-hairpin folding. *Proc. Natl. Acad. Sci. USA.* 100:12129–12134.

46. de Alba, E., M. Jimenez, and M. Rico. 1997. Turn residue sequence determines β-hairpin conformation in designed peptides. *J. Am. Chem. Soc.* 119:175–183.

47. Espinosa, J. F., F. A. Syud, and S. H. Gellman. 2002. Analysis of the factors that stabilize a designed two-stranded antiparallel β-sheet. *Protein Sci.* 11:1492–1505.

48. Rao, F., and A. Caflisch. 2004. The protein folding network. *J. Mol. Biol.* 342:299–306.