Article

Equilibrium Kinetic Network of the Villin Headpiece in Implicit Solvent

Weina Du[1] and Peter G. Bolhuis[1,*]
[1]van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT We applied the single-replica multiple-state transition-interface sampling method to elucidate the equilibrium kinetic network of the 35-residue-fragment (HP-35) villin headpiece in implicit water at room temperature. Starting from the native Protein Data Bank structure, nine (meta)stable states of the system were identified, from which the kinetic network was built by sampling pathways between these states. Application of transition path theory allowed analysis of the (un)folding mechanism. The resulting (un)folding rates agree well with experiments. This work demonstrates that high (un)folding barriers can now be studied.

INTRODUCTION

To elucidate and understand the detailed folding mechanism of proteins, it is desirable to have access to both experimental measurements and numerical simulation data. Up to now, only fast-folding proteins have fulfilled that criterion, since they fold sufficiently slowly for time-resolved experiments yet fast enough that their folding kinetics can be estimated by molecular dynamics simulations. One such fast-folding protein, the 35-residue subdomain of the villin headpiece (HP-35) (1,2) folds into a native state that is characterized by three α-helices packed together, thus forming a hydrophobic core (3,4) (see Fig. 1 for a visualization of the native structure). The experimental folding rate of the villin headpiece is on the order of a few microseconds (~0.5–5 μs^{-1}) (5–8). The experimental unfolding rate for wild-type HP-35 is much slower, specifically ~800 μs^{-1}, at 300 K (5). Thus, the stability of the native state with respect to the unfolded state is ~5–6 $k_B T$ (where k_B is Boltzmann's constant and T is temperature). Past experiments have revealed several important aspects of the villin headpiece folding/unfolding mechanism. Backbone mutagenesis (9) indicated that backbone hydrogen bonds mediate the helix formation from the transition state to the native state, and a folding nucleus found between helices II and III was predicted to accelerate folding. The finding that the transition state does not contain many native tertiary contacts and/or hydrogen-bonded helices, but may contain a hydrogen-bonded loop, is consistent with a relatively low folding free-energy barrier (10). Reiner et al. (11) used triplet-triplet energy transfer to monitor conformational fluctuations from equilibrium and found two conformational traps in the HP-35 unfolding process. They proposed that as it unfolds from the native structure, the protein encounters a relatively high barrier on the way to a major conformation where the C- and N-termini are proximate and then crosses a relatively low barrier on the way to a minor conformation where the two termini are farther apart.

The villin headpiece has been investigated extensively using molecular simulations (12–14) with both implicit (14–19) (including a study of a double mutant (19)) and explicit solvent (12,13,20–22). The first microsecond all-atom explicit-solvent molecular dynamics (MD) simulation of the villin headpiece took place in the late 1990s (20), and by only a few years later, the simulation length under the same condition had been extended 300-fold (12), by employing distributed computing. These long-time simulation trajectories clarified the folding mechanism to a large extent. A conformational search, after a quick (20-ns) molecular collapse, was proposed to be the rate-limiting step before the native state was finally reached (12). The folding rate was predicted to be ~5 μs. Freddolino and Schulten analyzed three folding trajectories from MD simulations (13). They found a folding time of 5.6–8.2 μs, during which the protein visited a long-lived intermediate state that has the native secondary structure, but with helix I flipped and rotated with respect to helix III. Folding to the native state was observed to proceed only after all three helices dissociate and reassociate with each other (13). A recent simulation of the Nle/Nle double mutant of the C-terminal fragment of the villin headpiece resulted in a folding rate of 2.8 μs at a melting temperature of 342 $\pm$ 5 K (23).

Similar to explicit-solvent simulations, implicit-solvent Monte Carlo simulations of the folding process by Yang et al. (14) showed early-stage collapse into a fairly compact structure, which remains compact until the native state is found, indicating the presence of well-formed secondary structure in the transition state. Later implicit-solvent simulations by Lei and Duan (17) proposed a well-defined two-stage folding process. Coming from the unfolded structure, the rate-limiting first stage leads to a folding intermediate

Submitted July 8, 2014, and accepted for publication November 14, 2014.
*Correspondence: p.g.bolhuis@uva.nl
Editor: Markus Deserno.

0006-3495/15/01/0368/11 $2.00
http://dx.doi.org/10.1016/j.bpj.2014.11.3476

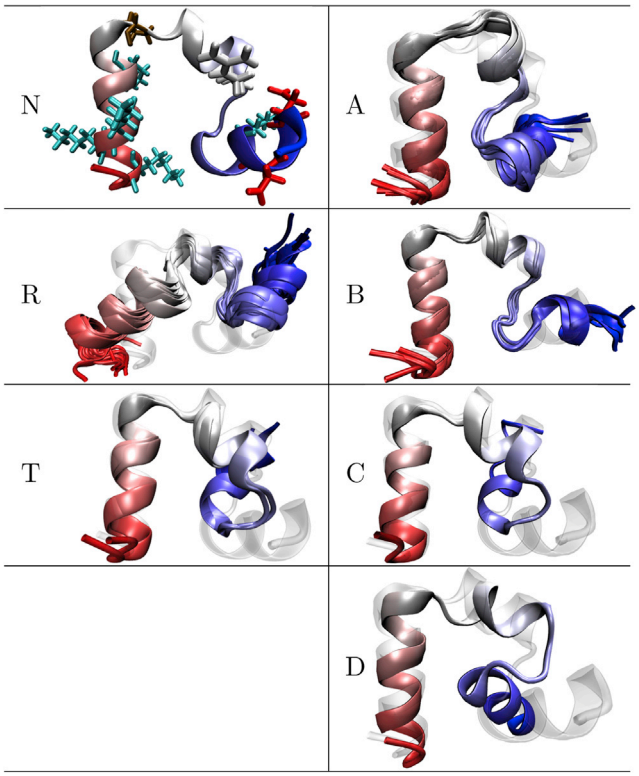

FIGURE 1 Structures for each state drawn with VMD (41). The protein backbone is shaded blue to red from the N-terminus to the C-terminus. For state N, several residues are highlighted: Lys in gray, Pro in brown, Arg in white, and Asp in red. States R and T are composed of multiple structures, with the native structure shown in transparent gray. States A and B are composed of multiple structures, whereas states C and D are single-structure states. To see this figure in color, go online.

where helices II and III form, which further facilitates the folding and docking of helix I in the second stage. Lei and Duan found on-pathway and off-pathway intermediates both with formed helices II and III. In contrast to the experiments by Reiner et al. (11), they found that the barrier separating the native state and the intermediate state is small, whereas that between the intermediate and the denatured states is large.

Most simulation approaches use either straightforward MD (23) or Markov-state modeling (24–26) to construct a (complex) transition network from a large but finite number of trajectories. However, when the trajectories spend most of their time in kinetic traps, these approaches can become inefficient. Parallel tempering or replica exchange, a popular simulation strategy to enhance sampling of protein conformational changes, efficiently generates a folding free-energy landscape but loses the direct connection to the kinetics. The path-sampling method (27,28) can alleviate these problems, since it samples unbiased rare trajectories between stable states and hence can sample transitions in which high barriers are encountered without losing any kinetic information.

In this study, we apply the recently developed single-replica multiple-state transition-interface sampling (SRMSTIS) (29,30) to HP-35 in implicit solvent. The aim of this work is to build up and analyze the equilibrium kinetic network for this protein and to show that we can overcome huge unfolding barriers (on the submillisecond timescale). For computational efficiency, we chose to work with a relatively small protein in implicit solvent. We note, however, that the method is fully applicable to explicit solvent (30). Starting with the native state, the SRMSTIS method builds up the equilibrium kinetic network by identifying metastable states and computing the kinetic-rate matrix between all these states. First, we performed a regular transition-interface sampling (TIS) of the native state to detect and define significant metastable states. Next, the SRMSTIS was applied to these states. During this sampling more metastable states were found, and these were added to the state set. A second SRMSTIS with nine states was applied to explore the kinetic network. Subsequent analysis using transition path theory (TPT) yielded new insight into the network, the unfolding mechanism, and the rate (31,32).

This article is organized as follows. In Materials and Methods, we introduce the molecular simulation and path-sampling conditions. In Results, we present the state identification along with the path sampling, followed by evaluation and analysis of the results. We end with concluding remarks.

MATERIALS AND METHODS

Molecular dynamics

All MD simulations were performed with the Gromacs-4.5.4 package (33–36). The experimental NMR structure (Protein Data Bank code 1YRF) (4) was used to initiate the simulation. We employed the CHARMM27 force field (37) with the multibody CMAP correction (38,39), which was parameterized to incorporate α-helix formation cooperativity in the force field (40) to deal with implicit-solvent systems. Protein configurations were visualized with VMD (41).

After putting the protein into a 6-nm cubic box, the energy was minimized by the steepest-descent method for 30 ps. In all simulations, the time step was 2 fs and the temperature was kept at 300 K by the v-rescale scheme (42). The nonbonded van der Waals cutoff radius was 1.4 nm, the LINCS algorithm constrained the bonds (43), and the Coulomb interaction cutoff was 4.8 nm. Frames were saved every 10 ps.

To represent solvation, we used the generalized Born surface area (GBSA) approach, which combines a GB electrostatic model with an approximation of the hydrophobic effect based on the solvent-accessible SA (44–49). Although the GBSA method can successfully identify native states of some short peptides with well-defined tertiary structure (50), in some studies, GBSA models predict different conformations compared with those from explicit solvent due to overly strong salt bridges and overpopulated α-helix (51).

Stable states were identified with the assistance of the k-centers cluster method (52,53).

Multiple-state transition interface sampling

The multiple-state TIS method samples the equilibrium kinetic network between predefined (meta)stable states and computes the corresponding rate

matrix. For each (meta)stable state I, the method requires a definition of a series of m nonintersecting interfaces based on an order parameter λ_{kI}, where each successive interface $1 \leq k \leq m$ delimits a volume in configuration space that encompasses the previous interface. The path ensemble belonging to each interface consists of all dynamical trajectories that start in the stable state and cross the interface, then return to the same state or end in any other state. The sampling of this interface-path ensemble is done via a Monte Carlo importance sampling by shooting off new trial trajectories from an existing pathway.

For this path sampling we use the one-way shooting algorithm (54). The stochastic one-way shooting move selects an initial point randomly from a previous accepted path and randomly decides whether to shoot forward or backward. For a forward move, an MD trajectory is initiated from the shooting point in the forward time direction until it reaches any stable-state definition. For a backward move, the momenta of all particles are reversed first, so that the integration effectively runs backward in time. The new trial path is constructed by joining the newly generated trajectory with the existing first half of the previous path (separated by the shooting point). This trial path can be accepted according to a metropolis Monte Carlo rule (for more details, see Du and Bolhuis (30)).

To identify the stable states, we used the root mean-square deviation (RMSD) of the backbone C_α atoms with respect to a reference structure. As the heads and tails of polypeptides are usually flexible, we use only C_α atoms from residue index 2–32. This order parameter, denoted $rmsd_{2-32}$, was used for both state and interface definitions. For instance, the native state S01 is defined by $rmsd_{2-32} < 0.13$ nm, i.e., any structure with a distance of <0.13 nm to the reference structure S01 is considered to be in the native state. Using the same order parameter, $\lambda \equiv rmsd_{2-32}$, a series of 10 nonintersecting interfaces were defined around S01, with increasing radii. Starting from 0.17 nm, neighboring interfaces are separated by a fixed interval of 0.04 nm. The TIS set of interfaces is thus defined by $\lambda = 0.17$, 0.21, 0.25, 0.29, 0.33, 0.37, 0.41, 0.45, 0.49, and 0.53 nm.

In addition, we employ the minus interface shooting to ensure faster decorrelation in the first interface ensembles (29,55). Every time the first interface of a state is visited, instead of performing a shooting move, we simply continue the MD trajectory from one end point until it crosses the first interface again and subsequently ends in some state. By cutting off the segment spent in the stable state(s), the last part of this trajectory forms a qualified trial path for the first interface ensemble. The average duration of MD trajectory between first crossings of the first interface also provides an estimate of the flux through the first interface (55,56). The MSTIS simulations were performed using a home-written Perl script. For more details, we refer the reader to Du and Bolhuis (30).

SRMSTIS simulations

Combining the MSTIS algorithm with a single-replica exchange method enables us to sample all interface ensembles for a single state in one path-sampling simulation (29). This SRMSTIS algorithm requires an optimal bias function, which can be constructed using a Wang-Landau approach, or, more efficiently, by a fixed bias, which is periodically updated. The optimal bias is equal to the crossing probability, i.e., the probability that a path will cross the interface provided that it left the metastable state at an earlier time. To improve convergence, an additional state swap move allows switching between different states, so that all defined states are sampled and paths decorrelate faster (29,30).

We performed SRMSTIS using multiple walkers. The ratio of trial interface exchanges to shooting moves is 9:1. Interface exchanging is governed by a fixed biasing function, denoted $\ln P(\lambda)$, which is updated after every 4500 interface exchange moves. State swapping allows paths to start in a different state, but it is only attempted in the outermost interface, for an accepted path that connects two different states and crosses the outermost interface of both these states. The state-swapping move is controlled by a Wang-Landau algorithm, with its own biasing function, which is updated every attempt (29). In this way, all states should eventually be sampled an equal amount. The initial (logarithmic) update factor is 0.0001, which is reduced by a factor of 2 once the sampling histograms of all states are equal within <10% of each other.

The path sampling can start from any interface of any state, with a qualified initial path. Ideally, with state swapping and interface exchanging, one single replica could accomplish equal sampling of all path ensembles given enough time. However, for efficiency, we employ multiple independent replicas, or walkers. These walkers store their results in a shared file, which is used to update bias functions for state swapping and interface exchanging. These bias functions govern the state-swapping and interface-exchanging behavior in all walkers (30).

Path-type analysis

The sampled path ensembles can be combined using a path-type analysis, as detailed in Du and Bolhuis (30). The rate-constant matrix is given by

$$k_{IJ} = \langle \phi_{1I} \rangle P_I(\lambda_{0J}|\lambda_{1I}), \tag{1}$$

where $\langle \phi_{1I} \rangle$ is the positive effective flux through the first interface λ_{1I}, and $P_I(\lambda_{0J}|\lambda_{1I})$ is the crossing probability for reaching state J (defined by λ_{0J}) from state I:

$$P_I(\lambda_{0J}|\lambda_{1I}) = \frac{\sum_{k=1}^{m} \tilde{n}_{IJ}(\lambda_{kI})}{\sum_{J \in M} \sum_{k=1}^{m} \tilde{n}_{IJ}(\lambda_{kI})}. \tag{2}$$

Here, the reweighted path-type number, $\tilde{n}_{IJ}(\lambda_{kI})$, is

$$\tilde{n}_{IJ}(\lambda_{kI}) = \overline{w}_k \sum_{i=1}^{m} n_{IJ}^{i}(\lambda_{kI}), \tag{3}$$

where $n_{IJ}^{i}(\lambda_{kI})$ is the path-type number for interface ensemble i, defined as the number of paths that start in state I, end in state J, and maximally reach interface λ_{kI}. The weights $\overline{w}_k = (\sum_{l=1}^{k} w_l^{-1})^{-1}$ are obtained from a weighted histogram analysis of the crossing probabilities for state I. See Du and Bolhuis (30) for more details.

RESULTS

Finding the stable-state set

Native-state TIS

The NMR (Protein Data Bank) structure after equilibration with MD, labeled as S01, defines the native state N. Starting from this state, we performed regular TIS in implicit solvent at 300 K. A sampling of the path ensembles at all 10 interfaces comprised around 4000 shooting moves in total. Of these, 314 long trajectories, including accepted paths >10 ns and rejected paths >5 ns, were collected for cluster analysis. Together, these 314 paths add up to ~6.4 μs, and given the frame interval of 10 ps, they contain 6.4×10^5 configurations. Using the k-center clustering method with $rmsd_{2-32}$ as the metric, all configurations were grouped into 2000 clusters with a near-equal radius of ~0.13 nm. Among these 2000 clusters, 55 clusters with >1000 members each comprise 79% of the total configurations. For each of these clusters, we extracted the

central structure based on the $rmsd_{2-32}$ metric. Further, we computed the $rmsd_{2-32}$ of all trajectories against the 55 extracted structures to check the clusters' stability. Here, a more strict criterion is applied, in which a structure is only qualified as a stable state when it is stable for >10 ns. Note that only paths >10 ns contribute in this step. This criterion yielded 42 stable structures.

A drawback of the used k-center clustering algorithm is that one well-populated state may be divided into two or more clusters, which are then kinetically very close to each other. By identifying correlations between groups of clusters in the trajectories, we lump these clusters together into a single state. Usually, such clusters are also conformationally similar, and in fact should have been lumped together in the clustering step. In this way, the 42 clusters were found to belong to just two states. The one containing 40 cluster centers is a major kinetic trap, whereas the other is a minor kinetic trap, with only two member clusters. Of the major trap, 28 clusters were very sparsely populated (<2%) and so were discarded. The remaining 12 structures were combined in the R state, whereas the two minor trap structures were considered to be part of the T state. In Table 1, the three stable-state definitions are listed in the first three rows, with the microstate (cluster) structures. Numbers in parentheses indicate the stable-state definition using the $rmsd_{2-32}$. For a visualization of their structures, see Fig. 1.

SRMSTIS with three states

Next, we performed SRMSTIS for the three states N, R, and T. Each state was defined by either a single structure (state N) or multiple structures (states R and T), as detailed in Table 1. Similar to the native-state TIS above, for states R and T, 10 interfaces were defined with an interval of 0.04 nm. When a state consists of multiple structures, e.g., state R or state T, we use the combination rule (57), which returns the minimum $rmsd_{2-32}$ distance of a trial structure to any of the state's member structures as the distance to the state.

In this first round of SRMSTIS, 20,000 paths were harvested for states N and R, whereas <3000 paths were sampled for state T, indicating that there are further kinetic traps in the vicinity of state T, which should have been defined as states. Using the same procedure as in the previous section, we collected 1218 trajectories >10 ns (accepted paths) or 5 ns (rejected paths) from all interface ensembles, adding up to 20 μs and comprising more than 2 million frames in total. Using k-center clustering, these configurations were grouped into 4000 microstates with an average radius of 0.13 nm. The RMSD centers of the 309 most populated clusters, each containing over 1000 members, were taken as microstates. Comparing the $rmsd_{2-32}$ metric of all trajectories against those 309 structures yielded 50 structures that trapped trajectories for >10 ns. Kinetic checking of those 50 structures resulted eventually in four additional macrostates, two of which are composed of five and four structures each, whereas the other two only contain a single structure each. The new macrostates were arbitrarily labeled A, B, C, and D, respectively. The definition of these macrostates by microstate structures and their cutoff in terms of the metric are also listed in Table 1. Their structures are shown in Fig. 1.

Besides these macrostates, we also added a denatured intermediate state, M, and an unfolded state, U, determined by deformation criteria of all three helices. State M has a denatured helix I and helix II, whereas helix III remains intact. For this state, the RMSD of helix I and helix II relative to the native state should be >0.30 nm and 0.25 nm, respectively, whereas the RMSD of helix III is within 0.15 nm of the native helix III. State U is defined by having all helices denatured, with an RMSD for the three helices of >0.30, 0.25, and 0.30 nm from the native helices I–III, respectively.

TABLE 1 Construction of all seven stable states by 26 structures

State	Structures					
N	1 (0.13)					
R	2 (0.13)	3 (0.13)	4 (0.13)	5 (0.13)	6(0.13)	7(0.13)
	8 (0.13)	9 (0.13)	10 (0.13)	11 (0.13)	12(0.13)	13(0.13)
T	14 (0.10)	15 (0.10)				
A	16 (0.10)	17 (0.10)	18 (0.10)	19 (0.10)	20(0.10)	
B	21 (0.10)	22(0.10)	23 (0.10)	24 (0.10)		
C	25 (0.10)					
D	26(0.10)					

The RMSD cutoff for each structure is given in parentheses.

Sampling the kinetic network

SRMSTIS with seven states

We carried out a SRMSTIS simulation for the system based on the extended state set $\mathbf{E}$ = (N, R, T, A, B, C, D, M, U). We allowed paths to start in any of the seven states (N, R, T, A, B, C, or D) and to end in any $X \in \mathbf{E}$. A set of 10 interfaces with a fixed interval of 0.04 nm was defined for each state, $X \in$ (N, R, T, A, B, C, and D). As described in the Materials and Methods section, we allowed for shooting moves, exchange moves, and state swapping.

Forty independent SRMSTIS simulations (walkers) were initiated, exchanging interfaces and states on the fly based on the biasing functions. In total, ~10,000 shooting moves were performed for each state (Table 2). The acceptance ratio for shooting moves is 54%, and those of interface exchange and state swap are 62% and 30%, respectively. The total aggregate simulation time of the sampled paths is ~200 μs. Convergence of the sampling bias functions is shown in Fig. 2. Fig. 3 gives an indication of how the 40 walkers randomly move through the interface space. (For a list of normalized path-length distributions for

TABLE 2 Number of trial shooting, interface exchanging moves, and state swaps for seven states

State	Shootings	Exchanges	Swaps
N	9309	84,567	118
R	10,011	89,190	238
T	10,016	87,757	178
A	8118	72,340	53
B	7575	67,244	145
C	9562	86,521	55
D	9278	84,736	220

ensembles from λ_1 to λ_{10} of all states, see Fig. 8 in the Appendix).

Based on a total of $\sim$ 64,000 trial shooting moves, we extract the flux and transition count matrices. Using path-type reweighting (see Du and Bolhuis (30)) we compute the crossing probability matrix, the rate matrix,

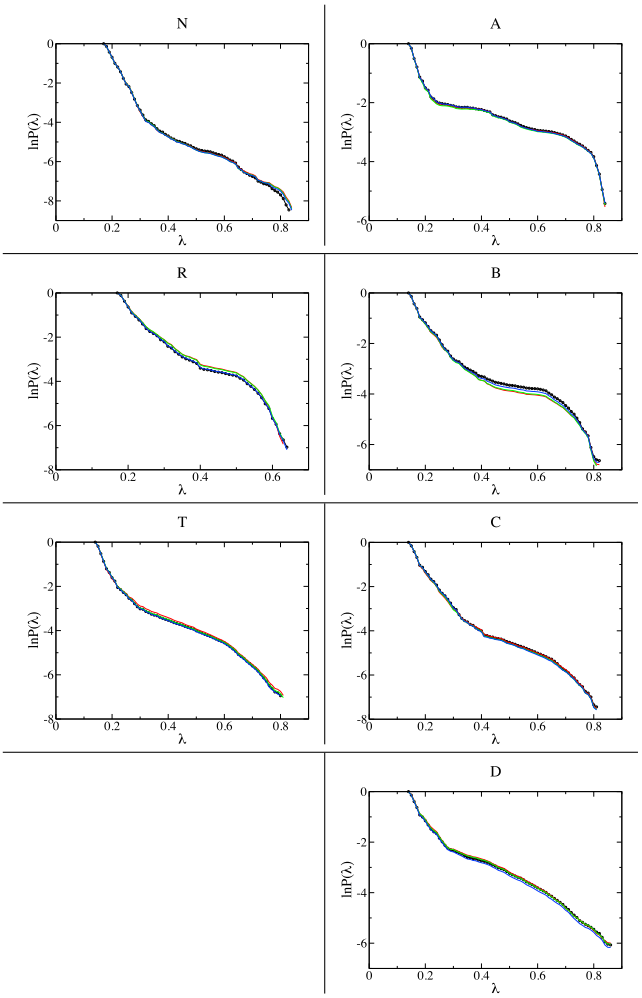

FIGURE 2 The last three biasing functions (in the order red, green, and blue, where blue is the last bias condition) for all sampled states, N, R, T, A, B, C and D, compared with the final crossing probability of the respective state (*black dotted line*). The near overlap of the curves implies convergence. To see this figure in color, go online.

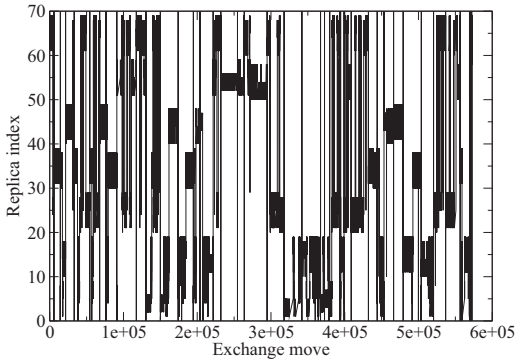

FIGURE 3 Replica index as a function of the number of exchange moves for all 40 walkers. On the *x* axis, the walkers are separated by vertical lines. On the *y* axis, each successive set of 10 replicas corresponds to a different state, with each replica in the set corresponding to λ_1–λ_{10}. States are in the order (lowest to highest) N, PN, SN, Mg, meta, and Pd.

and the transition time-length matrix (see Table 4). To check the reliability of the reweighting, we compare the crossing probability at each interface of a state as computed from path-type reweighting analysis with the results from ensemble-based crossing probability analysis. Deviations of crossing probabilities from the two analysis methods are small for all interfaces of the seven sampled states (see Table 7 in the Appendix).

During sampling, the denatured state M was sampled by many paths. Exemplary structures of state M are shown in Fig. 4. However, state U was not visited by any path. To gain more information about the unfolded state and build up the connection from U to any other states, we performed

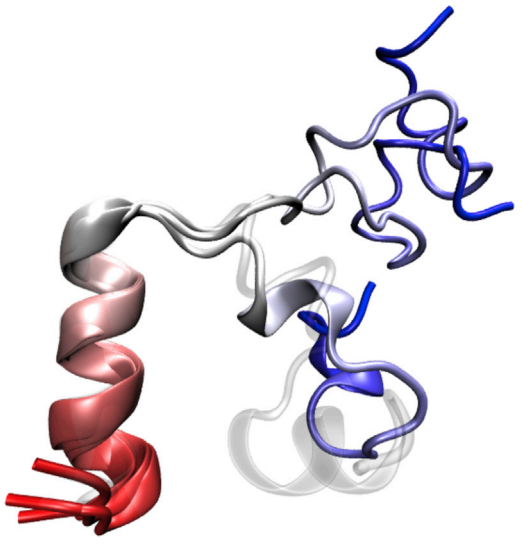

FIGURE 4 Three exemplary denatured structures in state M. The stable helix III is formed in all structures, but helices I and II are mostly deformed. The protein backbone runs blue to red from N-terminus to C-terminus. To see this figure in color, go online.

additional straightforward MD trajectories, which will be discussed in the next section.

MD trajectories start from fully unfolded structures

Eight 1-μs-long MD trajectories were started from randomly selected unfolded structures to extract more information from the unfolded region U and make a connection to the sampled network of states. As the experimental and theoretical folding time is ~5 μs, these MD trajectories should be able to sample the early stage of folding. Indeed, two out of the eight MD trajectories form helix III, whereas the remaining trajectories do not form helix III during the entire 1-μs trajectory. Two exemplary trapped structures in the unfolded region are shown in Fig. 5. Lysine residues form an obstacle to folding, especially in the formation of helix III. Helix III forms only slowly, but once formed it is stable and hard to deform. In contrast, helices I and II showed much more flexible and fast dynamics; they form and deform, merge into one helix or break up into two helices, and flip with respect to each other.

Flux at the first interface of sampled states

For the seven sampled states, the flux through the first interface, λ_1, was extracted directly from the inverse of the average length of the paths in the minus shooting move (55,56). For states M and U, the flux follows directly from straightforward MD simulation by keeping track of their mean first passage time (MFPT). We list the fluxes for all nine states in Table 3.

Analysis of the network

The transition probability and the rate matrices

SRMSTIS enables construction of the transition network among states N, R, T, A, B, C, D, M, and the MD trajectories initiated from the unfolded region yield transitions between M and U. By combining these two results, we construct the complete kinetic network for all nine states, as detailed in Du and Bolhuis (30). The transition matrix, the rate matrix

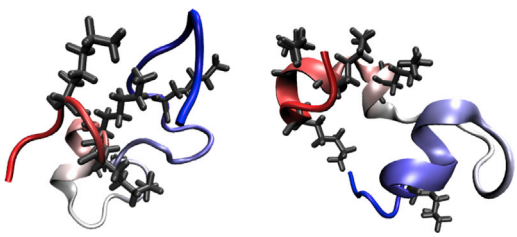

FIGURE 5 Two randomly selected metastable structures in the unfolded regime, visited by MD trajectories coming from the completely unfolded states. The protein backbone runs blue to red from N-terminus to C-terminus. Lysine is shown in gray. To see this figure in color, go online.

TABLE 3 Flux information at the first interface of each state

State	MFPT	Error	n_{path}	$\langle\phi_1\rangle$
N	0.1830	0.0191	2456	5.4648
R	0.3671	0.0075	1854	2.7241
T	0.3527	0.0681	2682	2.8354
A	1.3108	0.6954	978	0.7629
B	0.2454	0.0438	1146	4.0757
C	0.3653	0.2167	2462	2.7374
D	0.3125	0.2738	1490	3.1996
M	0.2223	0.1305	4226	4.4994
U	0.4947	0.2808	9389	2.0215

Columns denote, respectively, the state, the mean first passage time (MFPT) in ns, the error in the mfpt, the number of trajectories, and the average flux in units of ns^{-1}.

K, and the transition time matrix are shown in Table 4. In addition, we estimated the relative statistical error in the rates by assuming a normal distribution of the path-type numbers at the first interface. This statistical error estimate is mostly $<10\%$ and never higher than 50% for rarely sampled transitions.

These matrices indicate that, kinetically, A and B are close to N and R, whereas C and D are close to T. The M state acts as a bridge between all states and the unfolded state U. The reciprocal eigenvalues of the rate matrix yield the most important timescales in the kinetic network. These timescales in descending order are $\tau = \{\infty, 982.265, 378.568, 15.8943, 13.6639, 8.81709, 2.28238, 2.00341\}$ ns. The slowest timescale (apart from infinity, which is the equilibrium solution) corresponds to the overall folding process, and other timescales are fast processes related to transition between the metastable states. This slowest timescale corresponds to the single-exponential kinetics observed in time-resolved experiments. However, we also see a 378 ns timescale that has not been observed in experiments up to now, indicating the presence of at least one intermediate state.

Population of states

Applying $\mathbf{p}(t) = \mathbf{p}(0)\exp(\mathbf{K}t)$ for $t \to \infty$ for the rate matrix **K** yielded the stationary distribution for the nine states (see Table 5). Besides the native state, N, the only other significantly populated state is R, which accounts for ~30% of the overall population. In this state, all three helices are intact, but their relative positions are rearranged. Structurally, state R (defined by structures 2–13 (Fig. 1)) is very similar to the intermediate state identified by Lei and Duan (17). Those authors found that this state is an off-pathway intermediate that is not connected to N. However, in our calculations, the R state is kinetically very close to N, and carries a significant amount of unfolding flux (see section Transition-path theory analysis). The other, much less populated, states, T, A, B, C, and D (Fig. 1) have helix I flipped relative to helices II and III. This structural feature is very similar to the on-pathway intermediate state found in Lei and Duan (17)

TABLE 4 Transition probability, rate constant, transition time matrix, and relative error

	N	R	T	A	B	C	D	M	U
Crossing probability matrix									
N	9.96×10^{-1}	3.95×10^{-1}	8.96×10^{-6}	1.19×10^{-4}	1.22×10^{-4}	1.04×10^{-7}	1.84×10^{-6}	2.72×10^{-7}	
R	1.85×10^{-2}	9.81×10^{-1}	1.09×10^{-5}	2.58×10^{-4}	9.25×10^{-5}	4.90×10^{-7}	2.28×10^{-5}	1.49×10^{-6}	
T	4.94×10^{-4}	1.28×10^{-4}	9.93×10^{-1}	4.77×10^{-6}		1.27×10^{-3}	4.67×10^{-3}	1.70×10^{-4}	
A	6.44×10^{-2}	2.98×10^{-2}	4.67×10^{-5}	9.05×10^{-1}	9.66×10^{-5}		5.86×10^{-5}	4.51×10^{-4}	
B	1.05×10^{-1}	1.69×10^{-2}		1.53×10^{-4}	8.78×10^{-1}		2.62×10^{-5}	3.27×10^{-4}	
C	8.83×10^{-5}	8.87×10^{-5}	1.95×10^{-2}			9.78×10^{-1}	1.16×10^{-3}	1.01×10^{-3}	
D	6.08×10^{-4}	1.61×10^{-3}	2.80×10^{-2}	3.59×10^{-5}	1.01×10^{-5}	4.54×10^{-4}	9.68×10^{-1}	1.27×10^{-3}	
M	2.56×10^{-3}	2.99×10^{-3}	2.89×10^{-2}	7.87×10^{-3}	3.60×10^{-3}	1.12×10^{-2}	3.61×10^{-2}	9.03×10^{-1}	3.60×10^{-3}
U								5.32×10^{-4}	9.99×10^{-1}
Rate constant matrix (ns^{-1})									
N	—	2.16×10^{-2}	4.90×10^{-5}	6.52×10^{-4}	6.69×10^{-4}	5.70×10^{-7}	1.00×10^{-5}	1.49×10^{-6}	
R	5.03×10^{-2}	—	2.97×10^{-5}	7.03×10^{-4}	2.52×10^{-4}	1.34×10^{-6}	6.22×10^{-5}	4.05×10^{-6}	
T	1.40×10^{-3}	3.64×10^{-4}	—	1.35×10^{-5}		3.60×10^{-3}	1.32×10^{-2}	4.81×10^{-4}	
A	4.91×10^{-2}	2.27×10^{-2}	3.56×10^{-5}	—	7.37×10^{-5}		4.47×10^{-3}	3.44×10^{-4}	
B	4.27×10^{-1}	6.90×10^{-2}		6.25×10^{-4}	—		1.07×10^{-4}	1.33×10^{-3}	
C	2.42×10^{-4}	2.43×10^{-4}	5.34×10^{-2}			—	3.18×10^{-3}	2.76×10^{-3}	
D	1.94×10^{-3}	5.17×10^{-3}	8.96×10^{-2}	1.15×10^{-4}	3.24×10^{-5}	1.45×10^{-3}	—	4.06×10^{-3}	
M	1.15×10^{-2}	1.35×10^{-2}	1.30×10^{-1}	3.54×10^{-2}	1.62×10^{-2}	5.05×10^{-2}	1.63×10^{-1}	—	1.62×10^{-2}
U								1.08×10^{-3}	—
Transition time matrix (ns)									
N	—	46.36	20416.37	1533.86	1495.41	1753510.20	99588.61	671598.61	
R	19.88	—	33673.33	1422.81	3968.37	748853.16	16072.51	246901.23	
T	713.93	2745.38	—	73972.42		277.74	75.56	2080.43	
A	20.36	44.04	28082.82	—	13572.99		22370.08	2906.55	
B	2.34	14.49		1600.87	—		9367.68	749.82	
C	4136.34	4118.55	18.74			—	314.27	361.97	
D	514.45	193.58	11.16	8704.75	30905.90	688.23	—	246.29	
M	86.67	74.29	7.68	28.25	61.80	19.80	6.15	—	61.77
U								929.21	—
Estimated relative error in rates									
N	—	0.0123	0.1196	0.0445	0.0596	0.1544	0.0645	0.0692	
R	0.0149	—	0.2257	0.0526	0.0694	0.1949	0.0704	0.0494	
T	0.0695	0.1844	—	0.1709		0.0429	0.0202	0.0340	
A	0.0061	0.0098	0.2570	—	0.1429		0.1618	0.0226	
B	0.0063	0.0073		0.1211	—		0.2486	0.1027	
C	0.1510	0.1757	0.0115			—	0.0439	0.0184	
D	0.0741	0.0472	0.0090	0.1687	0.4431	0.0448	—	0.0248	
M	0.0378	0.0376	0.0071	0.0183	0.1504	0.0164	0.0102	—	0.0326
U								0.0755	—

Values were determined by path-type-reweighting analysis.

and to the folding intermediate proposed in Freddolino and Schulten (13), where the flipped helix I prevents the formation of the hydrophobic interactions present in the native state.

The relaxation of the population is shown in Fig. 6. From an initially completely folded state, there is a quick relaxation toward the R state within 50 ns. Relaxation to the other states is much slower, on the order of 1 μs (Fig. 7, *lower*).

Transition-path theory analysis

Using flux analysis (32) based on TPT (31,57), we computed the committor values for states R, T, A, B, C, and D from the overall transition from N to U (see Table 5). Based on the effective net flux matrix for the N $\rightarrow$ U transition (Table 6), the unfolding mechanisms can be summarized as a plot shown in Fig. 7. Note that direct flux transitions between (N, R, T, A, B, C, D) and U are rare. Although the rates for (N, R, T, A, B, C, D) $\rightarrow$ U are all zero in the rate matrix $\mathbf{K}$, they appear in the transition matrix, via the matrix exponent of $\mathbf{K}$. Because of symmetry, the net fluxes for the folding transition U $\rightarrow$ N are identical to the unfolding fluxes, but then, of course, reversed. Note also that the corresponding committor values are defined as $q_{\text{fold}} = 1 - q_{\text{unfold}}$.

State M transfers 84% net flux between the two sets, confirming its key role in the (un)folding process. Using TPT, we also computed the overall folding and unfolding mean FPTs (MFPTs) as 1.3 and 748 μs, respectively. We stress that to compute the rate of such a slow unfolding process using straightforward MD simulations would have taken orders of magnitude more computer time.

TABLE 5 Equilibrium populations and committors

State	Population	Committor
N	0.6719	0
R	0.2882	0.000036
T	0.0235	0.015
A	0.0089	0.00026
B	0.0011	0.00014
C	0.0016	0.016
D	0.0035	0.015
M	0.0001	0.049
U	0.0013	1

Corresponding to each state, the second column gives the population obtained from the rate matrix and the third column lists the committors for the N → U transition following from transition path theory analysis.

DISCUSSION

Using SRMSTIS, we computed the entire unfolding kinetics network of the villin headpiece. This resulted in identification of nine metastable states, and a 9 × 9 rate matrix. The experimental folding and unfolding times are 0.5–5 μs, and 800 μs, respectively, whereas our calculations yielded times of 1.3 and 748 μs, respectively, in very good agreement with experiment. Flux analysis shows that the unfolding goes via intermediates R or T (and to a much lesser extent A, B, C, and D), and a mandatory intermediate M. There are two prominent unfolding routes, N → T → M → U and N → R → D → M → U, observed from the flux plot (Fig. 7). The intermediate states have formed

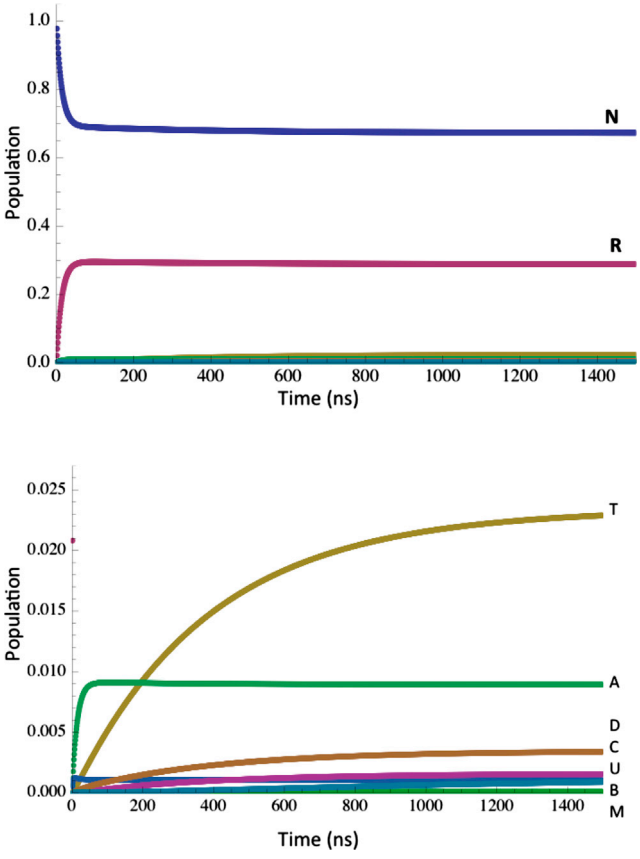

FIGURE 6 (*Upper*) Plot of the population relaxation, $\mathbf{P}(t) = \mathbf{p}(0)\exp(\mathbf{K}t)$, as a function of time t for all nine states in the villin headpiece system. At time $t = 0$, only the native state is populated. (*Lower*) Enlarged population versus time plot of the bottom region of the top figure, where the population relaxation of the less populated seven states can be seen. Lines in decreasing order of final population represent states T, A, D, C, U, B, and M, respectively. To see this figure in color, go online.

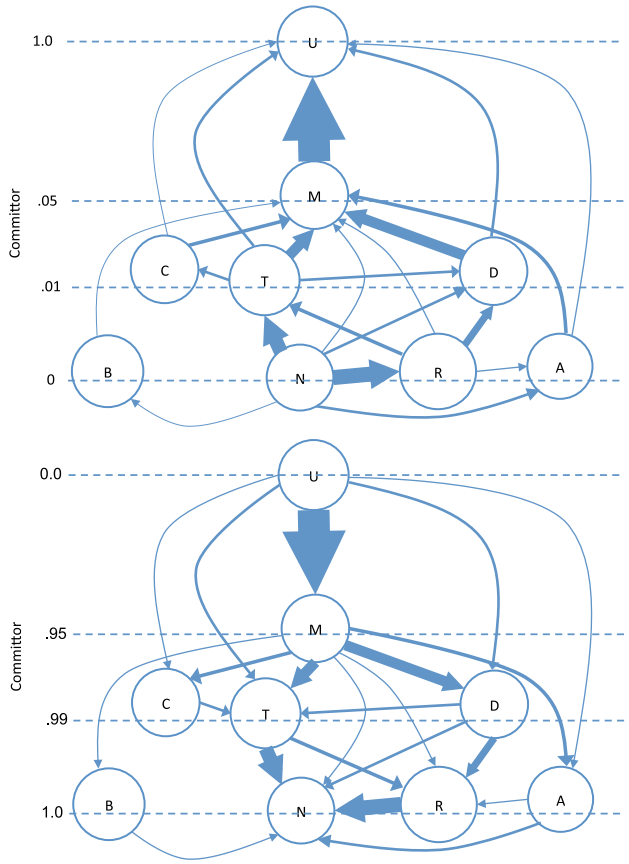

FIGURE 7 Villin headpiece flux for both unfolding (*upper*) and folding (*lower*). Numbers on the left indicate the committor probabilities of the corresponding levels. Thickness of the arrows indicates the N→U-related flux density. This figure only considers fluxes $>10^{-8}$ ns^{-1}. To see this figure in color, go online.

TABLE 6 Flux matrix of nine states for transition from N to U (in 10^{-8} ns^{-1})

	N	R	T	A	B	C	D	M	U
N		50.94	48.53	11.09	4.95	0.76	10.72	5.54	0.89
R			13.94	4.40	0.62	0.69	25.37	5.04	0.87
T						12.13	9.86	33.22	8.12
A			0.74			0.11	0.86	11.69	2.11
B			0.12	0.01		0.04	0.25	4.31	0.84
C								11.36	2.98
D						0.63		36.97	9.47
M									108.14
U									

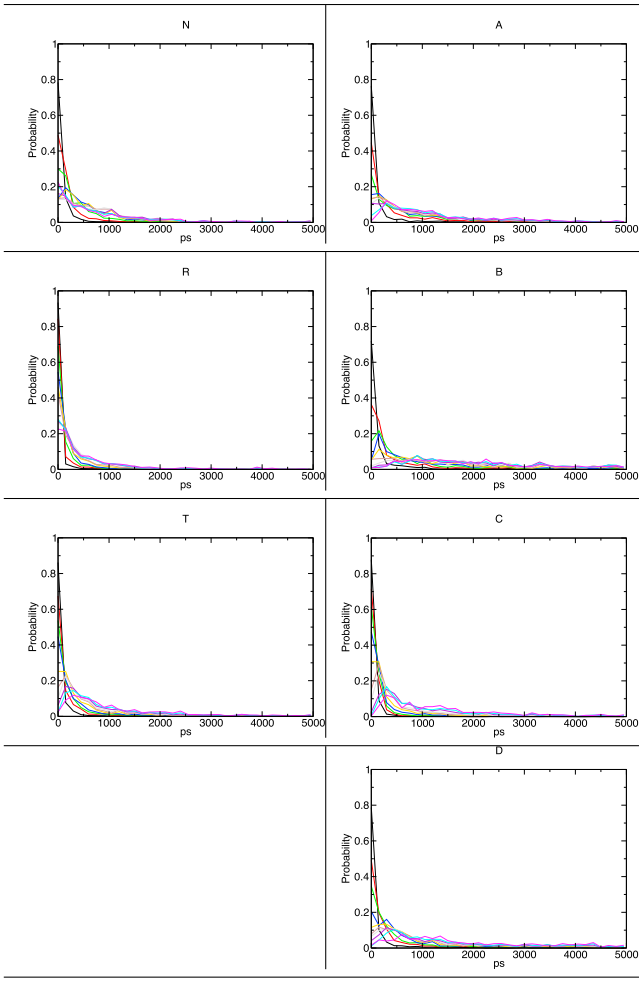

FIGURE 8 Normalized path length distributions for all sampled paths of the corresponding state. Ensembles from λ_1 to λ_{10} are shown with black, red, green, blue, yellow, brown, gray, violet, cyan, and magenta lines, respectively, at 150 ps resolution. To see this figure in color, go online.

helix II and helix III, which compared well with the previously detected on-pathway (states T, A, B, C, D) and off-pathway (state R) intermediates by implicit-solvent simulations (17). Moreover, we found a very fast N → R transition, a slower N → T transition, and extremely slow transitions R → U and T → U, in agreement with the conclusions of Lei and Duan (17) that the transition from native state to intermediates is easy, whereas that from the intermediates to the unfolded state is hard.

Both our SRMSTIS and straightforward MD simulations have proved that helix III is much more stable than helix I and helix II. In our SRMSTIS with seven structure-based states, the sampled paths witnessed the formation and deformation of helices I and II, but helix III was nearly always intact. In a similar way, MD simulations initiated from the unfolded state, as reported in the Results section, indicated that helix III was hard to form/deform, whereas helixes I and II were more flexible and exhibited (de)formation, helical merging, and bending more frequently. We thus defined the structures with intact helix III and denatured helix I and II as the intermediate state M, allowing it to connect the structure-based states to the fully unfolded region (with helix III deformed) as a transition.

Explicit-solvent simulations by Freddolino and Schulten (13) corroborate this finding. In all three MD folding trajectories, helix III was formed within the first 2 μs and remained stable ever after. The following 4–6 μs showed rearrangement of the relative positions of the three helices, and in one trajectory this was accompanied by formation and deformation of helices I and II before the native structure was reached. The different properties of helix III (stable) compared with the other two helices also has theoretical support, since helix III, as the longest helix in the villin headpiece, is composed of 10 residues, among which four are Lysine residues. Once formed, helix III is stabilized by those long residues due to the space constraints and/or electrostatic interactions.

Our unfolding flux calculation has highlighted the key role state M plays in (un)folding process (see Fig. 7). It acts as the mandatory intermediate state between the native and unfolded states. The importance of state M correlates with the properties of the three helices: helices I and II are flexible and experience multiple (de)formations, whereas helix III is the bottleneck of the (un)folding process.

CONCLUSIONS

To summarize, we have performed an extensive sampling of the kinetic network of the villin headpiece using SRMSTIS. Although the initialization phase is currently rather ad hoc and would benefit from automation, this work shows that this method can be used to sample complex folding networks, which would cost orders of magnitude longer simulation time with straightforward MD.

APPENDIX

In this appendix, we present additional information on the sampling. Fig. 8 shows the path-length distribution of all ensembles. Table 7 contains a comparison between the ensemble-based computation of the crossing probabilities and the path-reweighting analysis.

ACKNOWLEDGMENTS

This work is part of the research programme VICI 700.58.442, which is financed by the Netherlands Organization for Scientific Research (NWO).

REFERENCES

1. McKnight, C. J., D. S. Doering, …, P. S. Kim. 1996. A thermostable 35-residue subdomain within villin headpiece. *J. Mol. Biol.* 260: 126–134.

TABLE 7 Comparison of crossing probabilities from ensemble-based analysis and path-reweighting analysis for all states

State	Exchanges	Path type	Ensemble	State	Exchanges	Path type	Ensemble
		N				A	
0.17	17167	1.0000	1.0000	0.14	7293	1.0000	1.0000
0.21	15061	0.3475	0.3386	0.18	7351	0.3007	0.3268
0.25	15657	0.1226	0.1218	0.22	7906	0.1551	0.1787
0.29	12012	0.0390	0.0361	0.26	8544	0.1237	0.1497
0.33	11326	0.0178	0.0181	0.3	8361	0.1133	0.1363
0.37	9835	0.0107	0.0118	0.34	8544	0.1093	0.1342
0.41	9314	0.0077	0.0083	0.38	8495	0.1029	0.1255
0.45	9724	0.0061	0.0066	0.42	8530	0.0940	0.1148
0.49	9391	0.0047	0.0051	0.46	8332	0.0784	0.0959
0.53	10067	0.0040	0.0045	0.5	8132	0.0678	0.0821
		R				B	
0.17	12844	1.0000	1.0000	0.14	8961	1.0000	1.0000
0.21	11619	0.3843	0.3815	0.18	8098	0.3652	0.3937
0.25	11122	0.1857	0.1906	0.22	8137	0.2096	0.2315
0.29	11840	0.1143	0.1192	0.26	6684	0.1196	0.1225
0.33	11632	0.0658	0.0679	0.3	6529	0.0697	0.0761
0.37	11827	0.0437	0.0482	0.34	6913	0.0490	0.0561
0.41	13257	0.0298	0.0334	0.38	7407	0.0381	0.0424
0.45	12615	0.0265	0.0294	0.42	8709	0.0305	0.0357
0.49	12172	0.0228	0.0253	0.46	9094	0.0264	0.0311
0.53	11103	0.0154	0.0158	0.5	9577	0.0237	0.0287
		T				C	
0.14	18202	1.0000	1.0000	0.14	16,836	1.0000	1.0000
0.18	14816	0.2814	0.3002	0.18	18,287	0.3458	0.3717
0.22	14014	0.1203	0.1365	0.22	17,785	0.1773	0.1936
0.26	14041	0.0765	0.0844	0.26	16,224	0.0904	0.1014
0.3	13998	0.0476	0.0555	0.3	16,585	0.0483	0.0541
0.34	14664	0.0370	0.0427	0.34	14,525	0.0264	0.0309
0.38	14670	0.0308	0.0352	0.38	12,512	0.0188	0.0218
0.42	14615	0.0256	0.0291	0.42	12,572	0.0134	0.0158
0.46	14730	0.0217	0.0244	0.46	13,186	0.0119	0.0141
0.5	14909	0.0177	0.0200	0.5	13,648	0.0098	0.0116
						D	
				0.14	9926	1.0000	1.0000
				0.18	9290	0.3909	0.4161
				0.22	9888	0.2157	0.2318
				0.26	9915	0.1307	0.1345
				0.3	12,900	0.0947	0.1049
				0.34	12,383	0.0753	0.0835
				0.38	12,918	0.0673	0.0751
				0.42	13,513	0.0575	0.0657
				0.46	14,124	0.0465	0.0533
				0.5	14,500	0.0364	0.0419

Interface values of the state are shown in the first column. The second column shows the trial interface exchange moves. Columns 3 and 4 list crossing probabilities $P(\lambda_i|\lambda_1)$ at corresponding interfaces of the state from path-type-reweighting analysis and ensemble-based crossing probability analysis, respectively.

2. Frank, B. S., D. Vardar, …, C. J. McKnight. 2002. The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain. *Protein Sci.* 11:680–687.
3. McKnight, C. J., P. T. Matsudaira, and P. S. Kim. 1997. NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol.* 4:180–184.
4. Chiu, T. K., J. Kubelka, …, D. R. Davies. 2005. High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc. Natl. Acad. Sci. USA.* 102:7517–7522.
5. Kubelka, J., W. A. Eaton, and J. Hofrichter. 2003. Experimental tests of villin subdomain folding simulations. *J. Mol. Biol.* 329:625–630.
6. Wang, M., Y. Tang, …, D. P. Raleigh. 2003. Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale. *J. Am. Chem. Soc.* 125:6032–6033.
7. Brewer, S. H., D. M. Vu, …, R. B. Dyer. 2005. Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. *Proc. Natl. Acad. Sci. USA.* 102:16662–16667.
8. Brewer, S. H., B. Song, …, R. B. Dyer. 2007. Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy. *Biochemistry.* 46:3279–3285.
9. Bunagan, M. R., J. Gao, …, F. Gai. 2009. Probing the folding transition state structure of the villin headpiece subdomain via side chain and backbone mutagenesis. *J. Am. Chem. Soc.* 131:7470–7476.
10. Godoy-Ruiz, R., E. R. Henry, …, W. A. Eaton. 2008. Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J. Phys. Chem. B.* 112:5938–5949.

11. Reiner, A., P. Henklein, and T. Kiefhaber. 2010. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc. Natl. Acad. Sci. USA.* 107:4955–4960.
12. Zagrovic, B., C. D. Snow, …, 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.
13. Freddolino, P. L., and K. Schulten. 2009. Common structural transitions in explicit-solvent simulations of villin headpiece folding. *Biophys. J.* 97:2338–2347.
14. Yang, J. S., S. Wallin, and E. I. Shakhnovich. 2008. Universality and diversity of folding mechanics for three-helix bundle proteins. *Proc. Natl. Acad. Sci. USA.* 105:895–900.
15. Fernández, A., M. Y. Shen, …, K. F. Freed. 2003. Large-scale context in protein folding: villin headpiece. *Biochemistry.* 42:664–671.
16. Jang, S., E. Kim, …, Y. Pak. 2003. Ab initio folding of helix bundle proteins using molecular dynamics simulations. *J. Am. Chem. Soc.* 125:14841–14846.
17. Lei, H., and Y. Duan. 2007. Two-stage folding of HP-35 from ab initio simulations. *J. Mol. Biol.* 370:196–206.
18. Lei, H., C. Wu, …, Y. Duan. 2007. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 104:4925–4930.
19. Lei, H., X. Deng, …, Y. Duan. 2008. The fast-folding HP35 double mutant has a substantially reduced primary folding free energy barrier. *J. Chem. Phys.* 129:155104–155107.
20. Duan, Y., and P. A. Kollman. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science.* 282:740–744.
21. De Mori, G. M., G. Colombo, and C. Micheletti. 2005. Study of the villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics. *Proteins.* 58:459–471.
22. Tusell, J. R., and P. R. Callis. 2012. Simulations of tryptophan fluorescence dynamics during folding of the villin headpiece. *J. Phys. Chem. B.* 116:2586–2594.
23. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2011. How fast-folding proteins fold. *Science.* 334:517–520.
24. Chodera, J. D., N. Singhal, …, W. C. Swope. 2007. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126:155101.
25. Bowman, G. R., X. Huang, and V. S. Pande. 2009. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods.* 49:197–201.
26. Prinz, J.-H., H. Wu, …, F. Noé. 2011. Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134:174105.
27. Dellago, C., P. G. Bolhuis, …, D. Chandler. 1998. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* 108: 1964–1977.
28. Dellago, C., P. G. Bolhuis, and P. L. Geissler. 2002. Transition path sampling. *Adv. Chem. Phys.* 123:1–84.
29. Du, W. N., and P. G. Bolhuis. 2013. Adaptive single replica multiple state transition interface sampling. *J. Chem. Phys.* 139:044105.
30. Du, W., and P. G. Bolhuis. 2014. Sampling the equilibrium kinetic network of Trp-cage in explicit solvent. *J. Chem. Phys.* 140:195102.
31. Weinan, E., and E. Vanden-Eijnden. 2006. Towards a theory of transition paths. *J. Stat. Phys.* 123:503–523.
32. Noé, F., C. Schütte, …, T. R. Weikl. 2009. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA.* 106:19011–19016.
33. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.
34. Van Der Spoel, D., E. Lindahl, …, H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
35. 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.
36. Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.
37. MacKerell, A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
38. MacKerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* 126:698–699.
39. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.
40. Best, R. B., J. Mittal, …, A. D. MacKerell, Jr. 2012. Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of α-helix and β-hairpin formation. *Biophys. J.* 103:1045–1051.
41. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.
42. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
43. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
44. Still, W. C., A. Tempczyk, …, T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.
45. Scarsi, M., J. Apostolakis, and A. Caflisch. 1997. Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A.* 101:8098–8106.
46. Ghosh, A., C. S. Rapp, and R. A. Friesner. 1997. Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B.* 102:10983–10990.
47. Lee, M. S., J. F. R. Salsbury, and I. C. L. Brooks. 2002. Novel generalized Born methods. *J. Chem. Phys.* 116:10606–10614.
48. Lee, M. S., M. Feig, …, C. L. Brooks, 3rd. 2003. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* 24:1348–1356.
49. Im, W., M. S. Lee, and C. L. Brooks, 3rd. 2003. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* 24:1691–1702.
50. Ho, B. K., and K. A. Dill. 2006. Folding very short peptides using molecular dynamics. *PLOS Comput. Biol.* 2:e27.
51. Zhou, R. 2003. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins.* 53:148–161.
52. Gonzalez, T. 1985. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.* 38:293–306.
53. Dasgupta, S., and P. M. Long. 2005. Performance guarantees for hierarchical clustering. *J. Comput. Syst. Sci.* 70:555–569.
54. Bolhuis, P. G. 2003. Transtion path sampling on diffusive barriers. *J. Phys. Condens. Matter.* 15:S113–S120.
55. van Erp, T. S. 2007. Reaction rate calculation by parallel path swapping. *Phys. Rev. Lett.* 98:268301.
56. Bolhuis, P. G. 2008. Rare events via multiple reaction channels sampled by path replica exchange. *J. Chem. Phys.* 129:114108.
57. Du, W.-N., K. A. Marino, and P. G. Bolhuis. 2011. Multiple state transition interface sampling of alanine dipeptide in explicit solvent. *J. Chem. Phys.* 135:145102.