

Sampling the equilibrium kinetic network of Trp-cage in explicit solvent

Weina Du and Peter G. Bolhuis

Citation: *The Journal of Chemical Physics* **140**, 195102 (2014); doi: 10.1063/1.4874299

View online: <http://dx.doi.org/10.1063/1.4874299>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/140/19?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

[Enhanced sampling molecular dynamics simulation captures experimentally suggested intermediate and unfolded states in the folding pathway of Trp-cage miniprotein](#)

J. Chem. Phys. **137**, 125103 (2012); 10.1063/1.4754656

[Transition network based on equilibrium sampling: A new method for extracting kinetic information from Monte Carlo simulations of protein folding](#)

J. Chem. Phys. **135**, 235105 (2011); 10.1063/1.3670106

[Multiple state transition interface sampling of alanine dipeptide in explicit solvent](#)

J. Chem. Phys. **135**, 145102 (2011); 10.1063/1.3644344

[Kinetics and mechanism of the unfolding native-to-loop transition of Trp-cage in explicit solvent via optimized forward flux sampling simulations](#)

J. Chem. Phys. **133**, 105103 (2010); 10.1063/1.3474803

[Enhanced sampling and applications in protein folding in explicit solvent](#)

J. Chem. Phys. **132**, 244101 (2010); 10.1063/1.3435332



Re-register for Table of Content Alerts

Create a profile.



Sign up today!



Sampling the equilibrium kinetic network of Trp-cage in explicit solvent

Weina Du and Peter G. Bolhuis

*Van't Hoff Institute for Molecular Sciences, University of Amsterdam, PO Box 94157,
1090 GD Amsterdam, The Netherlands*

(Received 9 February 2014; accepted 20 April 2014; published online 21 May 2014)

We employed the single replica multiple state transition interface sampling (MSTIS) approach to sample the kinetic (un)folding network of Trp-cage mini-protein in explicit water. Cluster analysis yielded 14 important metastable states in the network. The MSTIS simulation thus resulted in a full 14×14 rate matrix. Analysis of the kinetic rate matrix indicates the presence of a near native intermediate state characterized by a fully formed alpha helix, a slightly disordered proline tail, a broken salt-bridge, and a rotated arginine residue. This intermediate was also found in recent IR experiments. Moreover, the predicted rate constants and timescales are in agreement with previous experiments and simulations. © 2014 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4874299>]

I. INTRODUCTION

Modeling the folding and unfolding of proteins in explicit solvent still proves a challenge for molecular dynamics (MD) simulations. The long times scales, often caused by large free energy barriers, render a straightforward MD simulation inefficient. While huge improvements have been made in the accessible MD time scale,¹ and in the analysis of MD trajectories in terms of Markov State Models,^{2–4} the fundamental issue that dynamics in complex system dominated by rare events is difficult to sample, remains. This problem is usually solved by biasing the sampling along a reaction coordinate. However, for protein folding these biasing methods often introduce spurious dynamics, lead to large hysteresis, and do not directly yield kinetic information. Transition path sampling (TPS)^{5–7} was developed to avoid the reaction coordinate problem, but in its basic form can only sample paths between two predefined states. The multiple state path sampling alleviates this limitation, and can, in principle, sample a whole network of transitions. The coupling with the transition interface sampling (TIS) methodology, namely, multiple state TIS (MSTIS),^{8,9} allows a direct calculation of the entire rate matrix for such a network resulting in a Markov State Model (MSM), but requires as many TIS calculations as there are states.¹⁰ This is in particular problematic for the replica exchange version of TIS, in which all interfaces have to be sampled simultaneously.^{11,12} Recently, we therefore developed a single replica version of MSTIS,¹³ which allows path sampling on a single interface that moves through the space of allowed interfaces, thus covering the entire network. The advantage is faster decorrelation of pathways, and almost complete decoupling of replicas. Here, we apply this methodology to the folding of a small model system: the Trp-cage mini protein.

Trp-cage, consisting of 20 residues, was designed to be a fast folding protein.¹⁴ Notwithstanding its small size, Trp-cage contains typical protein motifs, an α -helix (residue 2–8), a 3_{10} -helix (residue 11–14), and a polyproline II helix (residue 17–19), which form a hydrophobic core with the tryptophan buried in the center.

Many experimental methods have provided information concerning the folding mechanism of Trp-cage mini-protein. Qiu *et al.*¹⁵ applied laser temperature-jump spectroscopy and the measured folding rate of Trp-cage at room temperature is $k_f = (4.1 \mu\text{s})^{-1}$ with a relaxation time of $\tau = 3.1 \mu\text{s}$ (as $\tau^{-1} = k_f + k_u$, the unfolding rate $k_u = (12.7 \mu\text{s})^{-1}$). They reported a two-state folding process, and an unfolding midpoint of $T_f = 314$ K. Their results were later confirmed by Streicher and Makhatadze¹⁶ who carried out differential scanning calorimetry (DSC) and circular dichroism (CD) spectroscopy. However, other folding mechanisms have been proposed. Neuweiler *et al.*¹⁷ utilizing fluorescence correlation spectroscopy (FCS), proposed a more complex folding process with a collapsed molten globule-like intermediate as well as conformational flexibility of Trp-cage in the denatured state. This is in line with UV resonance Raman spectroscopy (UVRS) results carried out by Ahmed *et al.*,¹⁸ who proposed the presence of an intermediate with an intact α -helix, where the hydrophobic core is even more compact. Recent T-jump IR experiments¹⁹ established a biexponential relaxation time with $\tau_1 = 2.2 \mu\text{s}$ and $\tau_2 = 150$ ns at room temperature which suggested the additional presence of a native-like intermediate, characterized by a partly solvated proline helix.

Complementary to the experiments, molecular simulations have been carried out to gain atomistic insight in the folding process. Early implicit solvent all-atom simulation observed two-state folding and found a folding rate of $1.5\text{--}8.7 \mu\text{s}^{-1}$ at 300 K. Zhou²¹ employed replica exchange molecular dynamics (REMD) simulations in explicit solvent, which supported the two-state folding mechanism. An intermediate state was found where two hydrophobic cores are separated by an essential salt-bridge between residues Asp-9 and Arg-16 in the early stage of folding. Employing extensive all-atom MD simulations, Chowdhury *et al.*²² found a folding rate of $(\sim 3 \mu\text{s})^{-1}$ and proposed the packing of the central tryptophan as the rate-limiting step. Juraszek and Bolhuis²³ carried out all atom simulations in explicit solvent using TPS. They found Trp-cage folding paths that formed tertiary contacts and the salt bridge before formation of the helix, as well as paths that first formed the helix. These two folding routes

were also observed by Marinelli *et al.*²⁴ using bias exchange metadynamics simulations. These authors calculated that the folding process takes $\sim 2.3 \pm 0.7 \mu\text{s}$. Juraszek and Bolhuis²³ pointed out that the transition states of the rate-limiting steps are solvated native-like structures, and water expulsion is the last step for folding. They performed a full TIS calculation of the folding and unfolding process between intermediate loop state and the native state.²⁵ The unfolding rate turned out one order-of-magnitude higher than the measured experimental value, while the folding rate was in agreement with experiment. A later study of the transition employing optimized forward flux sampling (FFS) methods yielded results consistent with the TIS simulations.²⁶ The Shaw group published a 208 μs atomic-level MD applied on K8A mutant of the thermostable Trp-cage variant TC10b²⁷ and reported a folding time of 14 μs . Recently, Juraszek *et al.*²⁸ using the efficient partial path TIS method to study the (un)folding process, reported 4/6 μs and 5.4/3.3 μs , respectively, for the folding and unfolding time. Notwithstanding these good results, force field issues remained. Extensive REMD simulations by Day *et al.*²⁹ showed that the Amber99sb force field accurately described the equilibrium thermodynamics and folding transitions.

In this work, we employ the recently developed MSTIS approach to compute the equilibrium kinetic network of Trp cage. The aim is twofold: (1) to demonstrate the application of single replica MSTIS approach to relatively complex systems, (2) to establish a Markov state model describing the equilibrium kinetic network of Trp-cage, and provide further evidence for a (un)folding mechanism involving intermediate states. In this application, we construct a Markov state model for Trp-cage through several stages. The first stage is a regular TIS in which the native state is the only stable state. The resulting trajectories can become very long when the system is trapped in a metastable state. Analysis of the resulting accepted and rejected pathways leads to a definition of such metastable states. The resulting 14 metastable states are the basis for further TIS calculations. The second stage is application of the single replica MSTIS method with results into an estimate of the rate matrix. The matrix is completed using regular TIS for some lowly populated metastable states. The complete matrix is analyzed using Markov theory and Transition Path Theory (TPT).³

The paper is organized as follows. In Sec. II, we briefly review the MSTIS approach, and introduce a novel method to analyze the path ensemble for complex networks in which metastable states can fall inside the definition of the interfaces. In Sec. III, we present the initial TIS simulations and the cluster analysis that leads to the establishment of the 14 states, followed by the results of the single replica MSTIS simulations. We end with concluding remarks.

II. SIMULATION METHODS

A. Multiple state path sampling

1. Transition path sampling

For a brief overview of the path sampling methodology, we have to define a few concepts:

A **path** is a discretized dynamical trajectory $\mathbf{x}(L) \equiv \{x_0, x_1, x_2, \dots, x_L\}$ that consists of $L + 1$ times slices separated by time interval Δt . Each time slice contains positions and momenta of all particles in the system.

A **stable state** is a region in configuration space with a low free energy thus a high population, defined by one or a few selected order parameter(s) $\lambda(x) < \lambda_0$. State I is thus denoted by λ_{0I} .

An **interface** is a hypersurface defined by selected order parameter(s) $\lambda(x) = \lambda$ that surround the stable state. Interfaces are non-intersecting and the order parameter is usually monotonically increasing away from the stable state. λ_{sI} refers to sth interface of state I .

The effective positive **flux**, denoted $\langle \phi_{sI} \rangle$, is the average number of paths per unit time leaving state I to cross λ_{sI} , without first returning to I .

The **crossing probability**, denoted $P_I(\lambda_{kI}|\lambda_{sI})$, represents the conditional probability for a path to reach interface λ_{kI} before returning to I under the condition that it crosses λ_{sI} while coming directly from I . Note that it is allowed that the path returns to I after crossing λ_{sI} .

Developed for a two-state system, conventional TPS^{5,7} samples the transition path ensemble by performing a Monte Carlo random walk in the trajectory space. The **shooting algorithm**⁵ creates a trial path by selecting a random slice on the current path, changing its momenta slightly, and integrating the equations of motion forward and/or backward in time. The trial path is accepted when the path connects the two states. The original TPS scheme sampled trajectories of fixed duration. A slight variation is the flexible path length algorithm in which the integration is stopped upon touching a stable state.⁷ In this work, we use the one-way variant of the shooting algorithm, in which from the randomly chosen slice, without touching the momenta, the equations of motion are integrated either in the forward or the backward time direction. This improves the sampling efficiency, but requires a stochastic thermostat.³⁰

In complex systems, there are often multiple states, e.g., long lived metastable intermediates between the unfolded and a folded protein. When performing a two-state TPS in such a situation trial paths are likely to become kinetically trapped in these metastable states, thus dramatically increasing the average path length. The multiple state TPS method^{9,10} solves this problem by pre-defining all kinetic traps as stable states and by allowing all pathways that connect any two of these M states during sampling. As in the flexible length TPS the integration of the equation of motion is stopped once the path touches any defined state. The path ensemble consists thus of all possible paths between any two valid metastable states I and J with $I \neq J$. For this scheme, detail balance leads to the Metropolis acceptance rule,⁷

$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = h_I(x_0^{(n)})h_J(x_L^{(n)}) \min \left[1, \frac{L^{(o)}}{L^{(n)}} \right], \quad (1)$$

where the indicator function $h_I(x)$ is unity if x belongs to state I and zero otherwise. The labels (o) and (n) refer to the old and the new path, respectively.

2. Multiple state transition interface sampling

The TIS method allows an efficient evaluation of the rate constant and of the complete path ensemble by reweighting.³¹ The MSTIS, a direct generalization of TIS to more than two states, introduces a set of $m + 1$ interfaces λ_{sI} , $0 \leq s \leq m$ for every state I .^{9,10} In the usual convention, λ_{0I} denotes the boundary of I . For every state I and every interface λ_{sI} , MSTIS collects pathways that leave I and cross that interface before returning to I or going on to any other state J , by the shooting algorithm. The acceptance rule of paths in a MSTIS ensemble is

$$P_{acc}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = \tilde{h}_I^s(\mathbf{x}^{(n)}) \min \left[1, \frac{L^{(o)}}{L^{(n)}} \right], \quad (2)$$

where $\tilde{h}_I^s(\mathbf{x}^{(n)})$ is an indicator function that is unity if the path crosses the corresponding interface λ_{sI} and connects state I with any state J . Note that $J = I$ is now allowed. In addition, for I to I paths the **reversal** move changes the time direction of the path, by reversing the time slice order and negating all velocities. MSTIS enables the computation of all rate constants k_{IJ} for transitions between states I and J by evaluating^{8,9}

$$k_{IJ} = \langle \phi_{1I} \rangle P_I(\lambda_{mI} | \lambda_{1I}) P_I(\lambda_{0J} | \lambda_{mI}), \quad (3)$$

where $\langle \phi_{1I} \rangle$ is the flux through the first interface λ_{1I} of state I , which can be computed directly from straightforward MD.^{8,10} The second term $P_I(\lambda_{mI} | \lambda_{1I})$ is the crossing probability from λ_{1I} to λ_{mI} . This crossing probability cannot be computed directly, but needs to be constructed by combining the crossing probabilities $P_I(\lambda_{sI} | \lambda_I)$ for all m interfaces from λ_{1I} to λ_{mI} using, e.g., the weighted histogram analysis method (WHAM).³² The product of the first two terms gives the effective positive flux through the outermost interface λ_{mI} . The third term $P_I(\lambda_{0J} | \lambda_{mI})$ is the crossing probability for a path that crosses λ_{mI} while coming directly from I to reach state J before returning to I , and can be obtained from a MSTIS run of the outermost interface(s)

$$P_I(\lambda_{0J} | \lambda_{mI}) = \frac{n_{IJ}^m(\lambda_{mI})}{\sum_J n_{IJ}^m(\lambda_{mI})}, \quad (4)$$

where $n_{IJ}^m(\lambda_{mI})$ denotes the number of pathways starting in I , crossing λ_{mI} and ending in J . Note that this is an approximation that becomes only exact in the limit of infinite number of paths.

The full kinetics for a set of M states is given by a $M \times M$ rate constant matrix. Evaluation of the entire matrix requires n_M straightforward MD runs, $M \times (m - 1)$ separate interface sampling runs (where we assume all states have m interfaces), and a MSTIS run for all states at their outermost interfaces.

To enhance the sampling of the first interfaces, we employ the **minus-interface** sampling.^{11,12} The minus-interface move for λ_{1I} extends one (randomly chosen) endpoint of the path, into the stable state I until it crosses again the first interface λ_{1I} , and returns to any state. Because the trajectory typically spends a relatively long time in state I before it crosses the corresponding first interface again, it decorrelates from the previous path. The new trial path is identified as the part of the trajectory around the last positive (forward) crossing of the interface λ_{1I} , and can be always accepted. The minus

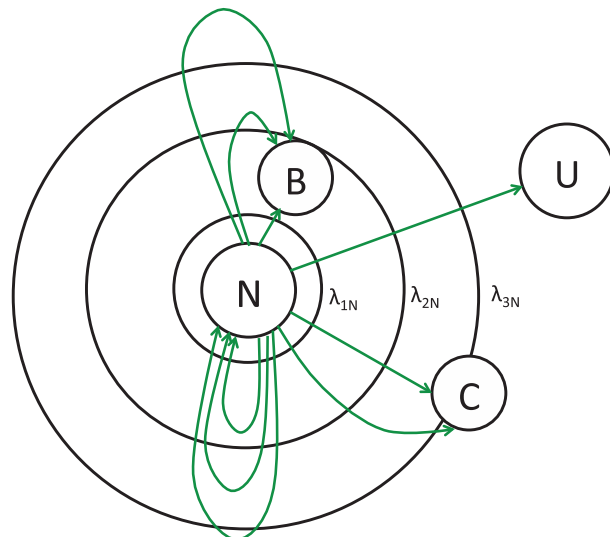


FIG. 1. Schematic representation of a system with four states: N , B , C , U . Only interfaces for state N are shown. B and C both fall within the outermost interface of N , while U is located outside of all interfaces. Green lines are examples of possible path types.

interface also provides the flux through the first interfaces $\langle \phi_{1I} \rangle = \langle \tau_I \rangle^{-1}$,¹¹ where τ is the time duration of the entire trajectory.

3. Path type reweighting analysis

Equation (3) assumes that all meta-stable states $J \neq I$ in the system are located outside the outermost interface of state I . While in theory the outermost interface can always be chosen to impose this situation, in practice it is hard to realize in a complex system, where (meta)stable states can sometimes be close to each other. See Fig. 1 for a graphical illustration. Moreover, if the outermost interface is chosen too close to state i , the sampling is not enhanced sufficiently. Therefore, it would be useful to allow overlap of the metastable states with the interfaces. However, when applying Eq. (3) many paths that avoid the outermost interface will be left out, leading to a wrong estimate of the rate. To account for these paths requires a modification of the way that transitions are counted and reweighted. First, instead of computing the rate from the crossing probability at the outermost interface we could measure it at the first interface

$$k_{IJ} = \langle \phi_{1I} \rangle P_I(\lambda_{0J} | \lambda_{1I}). \quad (5)$$

Of course, a direct computation of the second term is as difficult as the original rare event problem. However, just as it is possible to reweight the interface ensembles to get the crossing probability, it is possible to reweight the $P_I(\lambda_{0J} | \lambda_{1I})$ term. To do this, we split each interface ensemble i in sets of paths according to their **path type**, defined by its initial state I , its final state J , and the maximum interface λ_{kI} it has crossed. The number of paths observed in that set is denoted $n_{IJ}^i(\lambda_{kI})$ where the ensemble index is denoted by a superscript instead of an explicit condition to avoid confusion with the interface index of the path type. Note that the ensemble index i still refers to the condition that all paths have to cross λ_{iI} . The sum of the path type number over all $J \in M$ states, is simply the

number of paths that reach λ_{kl} . Adding these numbers over the interfaces $j \leq k \leq m$ results in the crossing probability for λ_{jl} when normalized with respect to λ_{il}

$$P_I(\lambda_{jl}|\lambda_{il}) = \frac{\sum_{k=j}^m \sum_{J \in M} n_{IJ}^i(\lambda_{kl})}{\sum_{k=i}^m \sum_{J \in M} n_{IJ}^i(\lambda_{kl})}. \quad (6)$$

Now, instead of reweighting the crossing probability, as was done previously, we can reweight the actual path type numbers $n_{IJ}^i(\lambda_{kl})$ themselves. The expression for this reweighted path type number is

$$\tilde{n}_{IJ}(\lambda_{kl}) = \bar{w}_k \sum_{i=1}^m n_{IJ}^i(\lambda_{kl}), \quad (7)$$

where $\bar{w}_k = (\sum_{l=1}^k w_l^{-1})^{-1}$ is the WHAM weight. Normalizing this number for all types gives the path type probability

$$\begin{aligned} \tilde{p}_{IJ}(\lambda_{kl}) &= \frac{\tilde{n}_{IJ}(\lambda_{kl})}{\sum_{J \in M} \sum_{k=1}^m \tilde{n}_{IJ}(\lambda_{kl})} \\ &= \frac{\bar{w}_k \sum_{i=1}^m n_{IJ}^i(\lambda_{kl})}{\sum_{J \in M} \sum_{k=1}^m \bar{w}_k \sum_{i=1}^m n_{IJ}^i(\lambda_{kl})}. \end{aligned} \quad (8)$$

The probability $P_I(\lambda_{0J}|\lambda_{1I})$ is now simply the sum over the path type probabilities for interface $1 \leq k \leq m$

$$P_I(\lambda_{0J}|\lambda_{1I}) = \sum_{k=1}^m \tilde{p}_{IJ}(\lambda_{kl}). \quad (9)$$

The rate constant matrix thus becomes

$$k_{IJ} = \langle \phi_{1I} \rangle P_I(\lambda_{0J}|\lambda_{1I}) = \langle \phi_{1I} \rangle \sum_{k=1}^m \tilde{p}_{IJ}(\lambda_{kl}). \quad (10)$$

It is also possible to compute the crossing probability for arbitrary λ_{jl} by again summing Eq. (8) over all states and all interfaces, yielding

$$\begin{aligned} P_I(\lambda_{jl}|\lambda_{1I}) &= \sum_{J \in M} \sum_{k=j}^m \tilde{p}_{IJ}(\lambda_{kl}) \\ &= \frac{\sum_{J \in M} \sum_{k=j}^m \tilde{n}_{IJ}(\lambda_{kl})}{\sum_{J \in M} \sum_{k=1}^m \tilde{n}_{IJ}(\lambda_{kl})}, \end{aligned} \quad (11)$$

which, for $i = 1$, is indeed identical to Eq. (6).

B. Single replica MSTIS

Evaluating the entire $M \times M$ rate matrix is quite an effort, as it requires on the order of $M \times m$ TIS simulations. Moreover, convergence of these simulations is not trivial, since each TIS ensemble is constrained to sample paths that cross the appropriate interface. A replica exchange scheme^{11,12} alleviates this convergence problem but is difficult to implement in parallel, since path lengths vary among the ensembles. Recently, we developed a novel approach in which a single replica randomly walks along the interface space.¹³ Interspersed with shooting and reversal moves, an **interface exchange** Monte Carlo move attempts to change the current interface to a neighboring interface. The standard metropolis

Monte Carlo acceptance rule dictates that the move can be accepted if the current path also crosses that interface. Because much more paths cross interfaces close to the stable state than interfaces further out on the barrier region, naturally the former are visited much more than the latter. Statistics is improved by imposing a bias which allows the sampling to visit all interfaces evenly. The natural probability to find a path at a certain interface, the so-called “density of paths” (DOP), is denoted by $g(\lambda)$. Performing a random walk with a probability proportional to $1/g(\lambda)$ results in a “flat” random walk across the interface space. The acceptance rule as required by detailed balance is then

$$P_{acc}(\mathbf{x}(L); \lambda_{iI} \rightarrow \lambda_{jI}) = \tilde{h}_I^j[\mathbf{x}(L)] \min \left[\frac{g(\lambda_{iI})}{g(\lambda_{jI})} \right], \quad (12)$$

where $g(\lambda_{ji})$ denotes the natural DOP at each interface λ_{ji} . Remarkably, the DOP (normalized with respect to the DOP at the first interface) is identical to the crossing probability $P_I(\lambda_{iI}|\lambda_{1I}) = g(\lambda_{iI})/g(\lambda_{1I})$. The algorithm requires knowledge of $g(\lambda_{iI})$, which can be obtained via either a Wang-Landau (WL) approach or by using the estimated crossing probability from a previous path sampling.¹³

1. Wang-Landau in single replica MSTIS

The WL algorithm^{33,34} attempts to achieve “flat sampling” by multiplying the biasing function $g(\lambda_{kl})$ with a factor f each time the algorithm visits interface k . Starting with all $g(\lambda_k) = 1$, the algorithm updates the bias until approximately all interfaces have been visited an equal amount of times, which is tracked by a visiting histogram $H(\lambda_k)$. A criterion could be, for instance, that the difference of any two $H(\lambda_i)$ should not be greater than 10% of the lowest value. When fulfilled, the WL factor is reduced to $f = \sqrt{f}$, and the visiting histogram is reset to zero. This procedure is repeated until convergence.

2. State swapping move

To allow the paths to start in a different state, we include an additional state swapping move when the current interface is the outermost interface. This move involves an attempt to reverse the paths time direction, and can be accepted if the current path connects two different states, I and J, and crosses both states’ outermost interfaces λ_{mI} and λ_{mJ} with the acceptance rule

$$\begin{aligned} P_{acc}(\mathbf{x}(L); \lambda_{mI} \rightarrow \lambda_{mJ}) \\ = \tilde{h}_I^m[\mathbf{x}(L)] \tilde{h}_J^m[\mathbf{x}(L)] \min \left[1, \frac{g(\lambda_{mI})}{g(\lambda_{mJ})} \right]. \end{aligned} \quad (13)$$

The time reversal of the path ensures that the path starts in J. This means that the state swapping move changes the ensemble from λ_{mI} to λ_{mJ} , with its corresponding DOPs. The WL criterion for flatness of the histogram $H(\lambda_{mI})$ is again set at 10%. When this criterion is fulfilled the increment factor f_s is reduced, and the sampling continues. Such a swapping move will eventually result in an equal sampling of outermost interfaces.

3. Fixed bias in single replica MSTIS

As the Wang-Landau approach in general converges slowly, one can adopt a fixed bias condition. The optimal bias is the crossing probability itself.¹³ The initial bias can be a crossing probability computed from a previous path sampling, based on experience or even an intelligent guess. An accurate initial bias condition is not required as the bias will be updated from time to time in the sampling procedure. While the update interval must be sufficiently long to allow the system to sample enough interfaces, it is not a sensitive parameter.

C. System setup and equilibration

We used the Gromacs package (version 4.0.5)^{35–38} for energy minimization and all molecular dynamics simulations. Employing the AMBER99sb force field,³⁹ the Trp-cage mini protein (PDB entry 1L2Y) was put into a rhombic dodecahedron box, solvated by 2798 TIP3P water molecules,⁴⁰ and neutralized by one chlorine ion. The energy was minimized by steepest descent followed by conjugate gradient optimization. To equilibrate the system, we performed position restricted MD simulation, followed by a molecular simulation at constant pressure of 1 atm. The box size was then set to the average value of 4.99596 nm, and all subsequent simulations were performed at constant volume. In all simulations, the time step was 2 fs and the temperature was kept 298 K by the v-rescale scheme.⁴¹ In all molecular simulation, the non-bonded van der Waals cutoff radius was 1.1 nm, the LINCS algorithm constrained bonds,⁴² and fast Particle-Mesh Ewald method treated long-distance electrostatic interactions.^{43,44}

D. Replica exchange molecular dynamics

We applied REMD⁴⁵ to sample the conformational space and extract reference configurational stable states to initiate the TIS runs. Two 40 ns REMD simulations were carried out starting, respectively, from the native structure and an unfolded structure. The unfolded structure was taken from a MD simulation at 500 K in which the protein became fully extended. In each of the two REMD simulations, 64 replicas are used with the temperatures exponentially distributed between 295 K and 564 K. In each replica, the velocity rescaling thermostat maintained the imposed temperature. Exchanging of all possible 64^2 replica pairs is attempted every 2 ps, and configurations are saved at the same frequency. The exchange of replica i and j is accepted with the usual rule $P_{acc}(i \rightarrow j) = \min[1, \exp((E_i - E_j)(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}))]$, where k_B is Boltzmann constant, T_i and T_j denote the temperatures of the two selected replicas, and E_i and E_j are, respectively, the potential energies at these two temperatures. We use the virtual move Monte Carlo move for better statistics.⁴⁶ The exchange move had an average acceptance ratio of 51%, based on first neighbor exchanges only (the total number of exchange trials is of course much larger). The two REMD simulations did not converge, thus did not lead to a reliable free energy landscape. (For a converged free energy landscape for this force field, we refer to Refs. 29 and 47.) Nevertheless, the REMD simulations can provide structural information in particular in the vicinity of the native state. Cluster analysis was applied to the

results of the folded REMD results in order to define states around the native state N.

E. Cluster analysis

We applied the Jarvis-Patrick clustering algorithm⁴⁸ to geometrically lump configurations into clusters. Two structures are identified as neighbors if their metric distance, here defined as RMSD of all heavy atoms of Trp-cage, is less than 0.15 nm. A neighbor list is constructed for each configuration from its M closest neighbors. The algorithm considers two configurations part of same cluster if they are in each others neighbor list, and have P common neighbors. We use $M = 15$ and $P = 3$ in our calculations.

In addition, we employed the k-centers clustering algorithm. K-centers clustering is faster than Jarvis Patrick and less memory demanding, and can handle large amounts of input. The method results in “equal-radius” clusters with variable population. There are four steps to follow.

- (1) Arbitrarily choose one configuration as the center of the first cluster.
- (2) Compute all RMSD metric distances between the existing center(s) to other configurations.
- (3) Assign the next cluster center, which is the configuration with farthest distance from the existing center(s), thus the total number of clusters increased by one.
- (4) Repeat step 2 and step 3 until the required number of clusters are generated. Distribute all configurations to all clusters, keep their radii relatively equal.

Cluster centers generated by k-centers clustering are not necessarily, and most of the time not, the geometric centers of the clusters, but only configurations initiated for the partition of the conformational space. To find the cluster centers for each cluster, one can perform a RMSD matrix calculation for all member configurations, and pick the configuration which has the minimum sum of RMSDs (with respect of all other members) to replace the old cluster center. The new centers, called RMSD centers, are more structurally representative of the clusters. This method combining k-centers and RMSD-center provides a better description of the conformational space.

While the k-centers clustering method is highly efficient it frequently suffers from state division problems. The division that k-centers clustering makes is rough, and cannot separate stable states from intermediate regions. Sparsely populated intermediate regions are also grouped into clusters (with low populations), while sometimes a well populated state is split into two or more clusters. To extract reliable stable states, we select only the clusters with high populations and check the dynamical trajectories to identify all structures (clusters) belonging to the same stable state.

III. RESULTS AND DISCUSSION

A. Defining the stable state set

1. Analysis of the folded REMD data

The REMD starting from the native structure yielded a collection of 20 000 structures at $T = 300$ K, which we used as

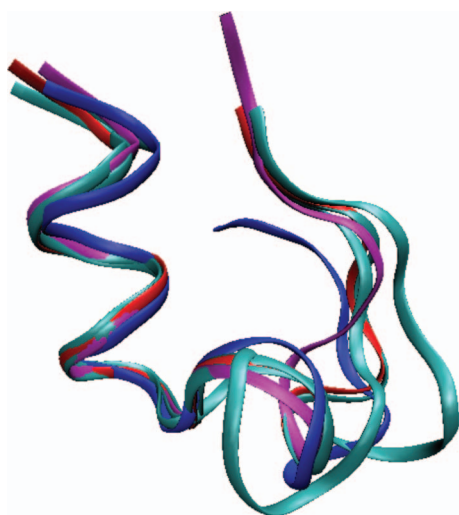


FIG. 2. Overlay of the six near native structures that are combined into one state definition for TIS. Red: native state N (S01). Blue: metastable neighbor (S06) frequently visited by paths leaving S01. Magenta: structure (S03) which turned out to be a key intermediate state in the (un)folding process. Cyan: the remaining structures S02, S04, S05.

input data for clustering. After Jarvis-Patrick cluster analysis, clusters were selected if they contained more than 50 members. To check stability of the cluster, a 100 ns MD run is performed starting from its central structure. If the trajectory at least stays for 4 ns within 0.15 nm of the center, a cluster is confirmed to be a metastable state. This analysis yielded five clusters, all close to the native state. MD runs of four clusters remained in their respective traps during the entire 100 ns simulation, while, the fifth cluster diffused away after 5 ns, suggesting a less stable structure. The most populated of these five clusters represents the native state, and its central structure is labeled as S01 (Figure 2). The other four clusters dwell in the neighborhood of S01, with their central structures labeled as S02, S03, S04, and S05. We detected a sixth (meta)stable state in the trial phase of TIS runs where paths from S01 frequently got trapped and extracted its central structure as S06. S06 is stable on 5 ns time scale, comparable with S05.

2. Escaping the native basin using TIS

Since all six structures identified are in the neighborhood of the native state, we view them as one stable state from which paths initiated to explore the configuration space using TIS. To combine the 6 structures S01 – S06 into one state N_6 , we apply the combination rule.¹⁰ The stable state indicator function $h_{N_6}[x]$ is unity if $x \in N_6 = S01 \cup S02 \cup S03 \cup S04 \cup S05 \cup S06$ and zero otherwise. The s th interface of combined state N_6 , is defined as $\{x : \min[\lambda_s(x)] = \lambda_s\}$, where the min function returns the minimum of its argument, $\lambda(x) = \{\lambda_1(x), \lambda_2(x), \lambda_3(x), \lambda_4(x), \lambda_5(x), \lambda_6(x)\}$, with $\lambda_{s1} \dots \lambda_{s6}$ the heavy atoms RMSD to the six centers. The combined interface is thus defined by the distance λ_s to the center of the nearest sub-state in the combined state.

We performed a regular TIS simulation initiated from the combined state. Interfaces were defined at $\lambda_{sN_6} = 0.20, 0.25, 0.30, 0.35$, and 0.45 nm, with a TIS simulation carried out for each one. In this TIS, we sample paths that start from N_6 , cross the combined interface and come back to N_6 . For interface λ_{sN_6} , at least one time slice a in the path should obey $\min[\lambda(a)] > \lambda_{sN_6}$. Note that since we do not have other states yet, paths have to return to N_6 . Also note that we have to include all six structures in the stable state definition, since paths can begin and end in any of S01 – S06. All TIS calculations taken together resulted in totally 6454 pathways, 2099 of which are accepted. The acceptance ratio of path ensembles decreased from 40% to 25% as the interface value increases from 0.20 nm to 0.45 nm. The total length of all paths together add up to $\sim 10 \mu\text{s}$.

3. Analysis of the TIS data

The ensemble of paths escaped the native basin and explored the Trp cage conformational space. We use this ensemble to identify a collection of (meta)stable states from which we will perform MSTIS (see Sec. III B). As metastable states are kinetic traps, we detect them by considering only the paths (both accepted and rejected) that are longer than 4 ns. Those paths together contain 963 309 frames, which are grouped into 10 000 clusters by k-centers clustering^{49,50} using RMSD of heavy atoms as distance. 125 individual clusters were found containing more than 1000 members, together accounting for 44.6% of overall population. For each of the 125 clusters we extracted the central structure, the structure with the minimum sum of RMSD with respect to all members of that cluster, as a candidate structure for further testing. Stability was checked by computing the RMSD of all trajectories against all 125 structures. A structure was considered stable if at least one path dwelled longer than 4 ns in it. Since some of the remaining 52 structures might actually belong to the same state, we carefully checked the trajectory behavior with respect to each structure. Structures with closely correlated kinetics belong to a same state and were combined. For example, if trajectories always enter and leave structure 12, 13, and 14 (almost) simultaneously, we can define a state Pd that is composed by the combination of three structures. This test reduced the 58 structures (which include the 6 structures of the initial state) to a set of 32 states (see Table I).

An additional 33rd state is the unfolded state U , which was not defined by structures due to the large conformational entropy of the unfolded region. Instead, we defined the U state as having a RMSD of the helix compared with an ideal helix over 0.25 nm and a RMSD of all heavy atoms with respect to the native structure larger than 0.55 nm. Thus, the conformational space of Trp-cage system contains (at least) a set of 33 metastable states ($N, PN, SN, Mg, meta, Pd, LNA, LNB, LNC, LSNA, LSNB, LMg, LmetaA, LmetaB, A, B, C, D, LE, F, G, LG, SI, PI, H, O, Q, R, T, La, Lb, W, U$). As building up a complete Markov state model for Trp-cage using MSTIS is a tremendous effort, we further reduced the metastable state set to $\mathbf{M} = (N, PN, SN, Mg, meta, Pd, LN, LSN, Lm, Lo, I, W, other, U)$ by combining similar and very lowly populated states among the states indicated by TIS analysis. This

TABLE I. The definition of 13 states by combining 32 microstates constructed by 58 structures. The stable state definitions is given between brackets (in nm).

State	Microstate	Structures
N	N	1(0.13)
PN	PN	2(0.10)
SN	SN	3(0.14)
Mg	Mg	5(0.12)
meta	meta	6(0.15)
Pd	Pd	12(0.14), 13(0.14), 14(0.14)
LN	LNA	4(0.12)
	LNB	17(0.14)
	LNC	37(0.15), 38(0.15)
LSN	LSNA	9(0.14), 10(0.14)
	LSNB	21(0.12), 22(0.12), 23(0.12)
Lm	LMg	39(0.10)
	LmetaA	25(0.12)
	LmetaB	40(0.12), 41(0.12)
Lo	LE	18(0.15), 19(0.15), 20(0.15)
	LG	27(0.14), 28(0.14)
	La	51(0.10)
	Lb	43(0.16), 44(0.16), 45(0.16), 46(0.16), 47(0.16), 48(0.16), 49(0.16), 50(0.16)
I	SI	29(0.18), 30(0.18), 31(0.18), 32(0.18)
	PI	33(0.15)
W	W	57(0.30), 58(0.30)
Other	A	7(0.13)
	B	8(0.12)
	C	11(0.13)
	D	15(0.12), 16(0.12)
	F	24(0.10)
	G	26(0.12)
	H	34(0.15), 35(0.15), 36(0.15)
	O	42(0.12)
	Q	52(0.14)
	R	53(0.13)
	T	54(0.15), 55(0.13), 56(0.14)

14-state set is used in all subsequent calculations and analyses (see Table I). Structure S01 is relabeled the native state *N*. State *PN* (S02) is a metastable state, very close to *N*, but with a more extended hydrophobic core and 3_{10} helix caused by Ser13 bond rotation. State *SN* (S03), has an intact α helix and 3_{10} helix, but a broken Asp-Arg salt bridge and a slightly sifted and rotated proline tail with respect to *N*. Compared to the native state, state *Pd* has a deformed 3_{10} helix, state *Mg* has a twisted relative position of α helix and 3_{10} helix, and in state *meta* the proline tail is rotated halfway towards the back of the molecule. States with “*L*” in their labels are composed of structures with a head-deformed α helix and intact hydrophobic core, among which “*LN*” has the same backbone with state *N*, “*LSN*” has the same backbone with state *SN*, “*Lm*” has members with backbone of either *Mg* or *meta*, and members of “*Lo*” mainly have deformed 3_{10} helix. In state *I* the α helix is intact but the proline tail rotated

away, exposing the hydrophobic core to the solvent. In state *W*, the first half of α helix is deformed but the hydrophobic core is still intact. The remaining states were combined to a state labeled “*other*.” Structures of all 32 microstates and the composition of 13 states by the 32 microstates are shown in Fig 3. Note that *LN*, *LSN*, *Lm*, *Lo*, *I*, and *other* are combined states, i.e., each is composed of more than one microstate.

B. Sampling the equilibrium kinetic network

1. Single replica multiple state transition interface sampling simulations

We start with exploring the local network around *N*. We employed the single replica MSTIS for the first six states (*N*, *PN*, *SN*, *Mg*, *meta*, *Pd*), among which we allowed state swapping, i.e., not only interfaces of a same state, but also the outermost interfaces of two different states are allowed to exchange. State swap and interface exchange accelerate path decorrelation and yield equal sampling of interfaces and states that are differently populated naturally.

The one-way shooting algorithm with a uniform random choice of shooting point generated and accepted trial paths (see Sec. II and Ref. 30). The number of interface exchange moves to number of shooting moves is 9:1. State swap moves are attempted with a ratio of 1:1 to shooting moves. When an exchange with the first interface is attempted, the minus interface move takes place. The exchange move is governed by a fixed (logarithmic) bias $\ln DOP$, given by the crossing probability based on the sampled paths. This crossing probability and thus the bias function is updated after every 4500 exchange moves, i.e., around 500 shooting moves. The state swap move is achieved by an interface exchange move between two outermost interfaces of two states by employing the Wang-Landau algorithm, initiated with a (logarithmic) increment of 0.001 on each visit. For this simulation, 5 interfaces were defined for each state, evenly distributed with an interval of 0.04 nm. The interface values are shown in the first column of Table II. The first interface location is chosen to yield a flux on the order of nanoseconds in the minus interface move. For each state 3 walkers were initiated. Via interface exchanging and state swapping, all 18 walkers contribute to and are governed by the same bias function. The acceptance ratio for shooting was 51%, for interface exchange 27%, and for state swapping 11%. The final analysis is based on 12 048, 6401, 9820, 7372, 8086, and 4887 paths for the respective states. The convergence of the sampling bias is shown in Fig. 9 of the Appendix. Fig. 4 shows how the 18 different walkers sample the interfaces. Clearly, state swapping occurs less frequent than interface exchanging, as expected.

Besides the six states intensively sampled as described above, the Trp-cage system contains 8 other metastable states. Among these, state *I* has a polypoline tail that points away and exposes the hydrophobic core, state *W* has the α -helix deformed and hydrophobic core preserved and state *U* is a broad region contains a big variety of unfolded state structures. In order to derive a full network, we performed three additional sets SRMSTIS for each of these three states. The

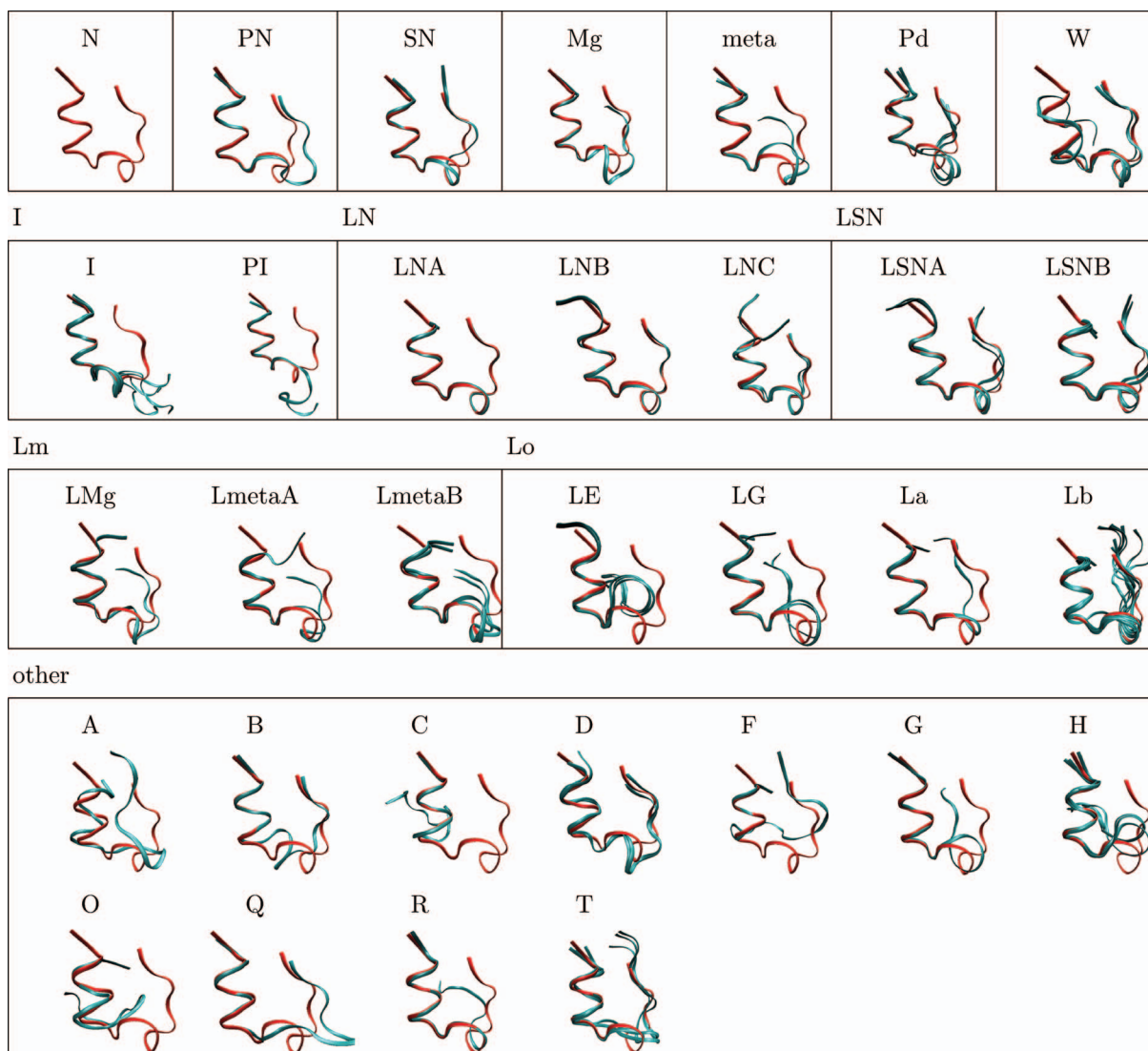


FIG. 3. Structures of the 13 stable states, composed from 32 microstates based on 55 structures. The native structure is shown in red in each picture (except N) for comparison.

three states are independent from the local network and show no fast inter-conversion. Therefore, we deactivated the state swapping move, and only allowed interface exchanging. The one-way shooting algorithm was employed. For state *I* and *W*, we kept the uniform random selection of shooting point, while for state *U*, we employed a Gaussian bias (with a standard deviation of 0.1 nm from the interface) to select shooting points.²⁵ The acceptance ratio for shooting was 55%, for interface exchange 43%. For state *I*, *W*, and *U*, respectively, 4333, 4803, and 2665 paths were collected. The converged bias condition $\ln\text{DOP}$, i.e., the crossing probability of all 3 states are shown in Fig. 9 of the Appendix.

The 9 TIS ensembles were subjected to the path analysis described in Sec. II A 3. Applying Eq. (11) yields the crossing probabilities $P(\lambda_{IX}|\lambda_{IX})$ (see Table II). We can compare these crossing probabilities with the results of the (regular) ensemble based analysis.¹³ Table II indicates that both methods perform comparable.

The remaining five states *LN*, *LSN*, *Lm*, *Lo*, and *other* turn out not very stable, nor very populated. We nevertheless sam-

pled the path ensembles coming out of these states by TIS, to obtain a more complete picture. It turned out that sampling on the first interface (λ_1) defined at 0.03 nm larger than the state definition (see Table I), including the minus move provided sufficient transitions. For states *LN*, *LSN*, *Lm*, *Lo*, and *other*, respectively, 447, 759, 97, 307, and 902 paths were harvested, which can be used to construct the complete matrix. While some MSTIS paths were a few ns long, the majority had a duration of less than a 1 ns. The transitions between states can take longer (see Table VI in the Appendix). In all, the aggregate simulation time of the (MS)TIS simulations is 153.8 μs .

Fluxes of all sampled states at their respective first interfaces are listed in Table III, extracted from the implemented minus interface sampling method (see Sec. II).

2. The rate matrix

For each state, the path type analysis yields the number of paths that end in any of the 14 states. To turn this into

TABLE II. Crossing probabilities for each state, as given by path type analysis and by the regular method. Columns denote, respectively, the interface value, the number of exchanging moves, the crossing probability from path type analysis, and the crossing probability from regular analysis.

λ	Exchanges	Path type	Ensemble based
N			
0.16	36 692	1.0000	1.0000
0.2	34 590	0.1001	0.1136
0.24	32 661	0.0255	0.0304
0.28	31 282	0.0073	0.0091
0.32	36 493	0.0024	0.0031
PN			
0.13	17 933	1.0000	1.0000
0.17	15 512	0.4748	0.4894
0.21	13 605	0.1631	0.1764
0.25	12 937	0.0432	0.0472
0.29	6965	0.0168	0.0211
SN			
0.17	39 677	1.0000	1.0000
0.21	33 895	0.2218	0.2368
0.25	33 938	0.0544	0.0636
0.29	27 529	0.0146	0.0176
0.33	26 796	0.0048	0.0061
Mg			
0.15	17 552	1.0000	1.0000
0.19	18 835	0.4272	0.4696
0.23	16 985	0.1347	0.1560
0.27	16 975	0.0420	0.0501
0.31	10 773	0.0093	0.0122
meta			
0.18	17 364	1.0000	1.0000
0.22	19 727	0.3661	0.3771
0.26	20 589	0.1791	0.1936
0.3	17 414	0.0474	0.0547
0.34	11 154	0.0101	0.0134
Pd			
0.16	17 070	1.0000	1.0000
0.2	10 598	0.2834	0.3033
0.24	10 146	0.1051	0.1144
0.28	9148	0.0498	0.0572
0.32	8511	0.0268	0.0323
I			
0.204	13 499	1.0000	1.0000
0.244	13 500	0.4955	0.5291
0.284	13 555	0.3520	0.3819
0.324	12 227	0.2648	0.2934
0.364	13 804	0.2199	0.2451
W			
0.33	13 326	1.0000	1.0000
0.37	13 091	0.3039	0.3306
0.41	12 468	0.1538	0.1655
0.45	12 542	0.0742	0.0782
0.49	11 503	0.0388	0.0424
U			
0.08	16 365	1.0000	1.0000
0.12	7097	0.1488	0.1627
0.16	6223	0.0669	0.0753
0.2	4467	0.0286	0.0317
0.24	5065	0.0172	0.0178

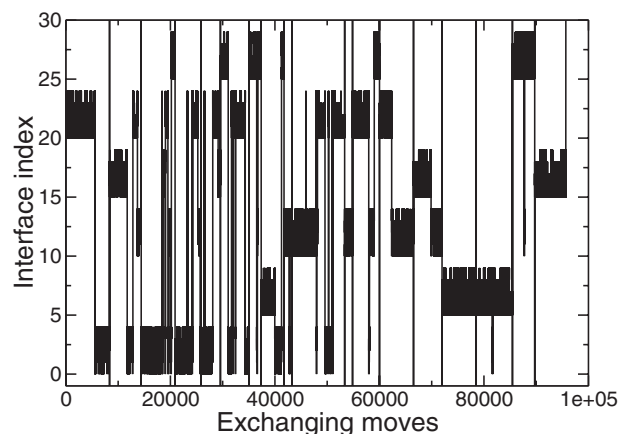


FIG. 4. The replica identity as a function of exchange move index, for each of the 18 walkers. Walkers are separated by vertical lines. The replica index corresponds to λ_1 to λ_5 for state N, PN, SN, Mg, meta, and Pd, respectively.

a rate matrix, we symmetrized the path count matrices for each state by setting $n_{IJ} = n_{JI}$, where $n_{IJ} \equiv \tilde{n}_{IJ}(\lambda_{IJ})$ (see Eq. (7)), as in an unbiased ensemble each IJ path should be as likely as its reversed JI path. For the i th state, this resulted in a path count matrix with only a non-zero row i and a non-zero column i . All 14 matrices were joined using WHAM³² which resulted in a 14×14 transition matrix (see Table VI in the Appendix). Combining this matrix with the flux, and normalizing using $k_{IJ} = \langle \phi_I \rangle P_I(\lambda_{0I} | \lambda_{1I}) = \langle \phi_I \rangle n_{IJ} / \sum_J n_{IJ}$ (cf. Sec. II A 3), yields the final rate matrix, as shown in Table IV. This table also reports the mean first passage times $\tau = k^{-1}$ for all transitions. Numbers corresponding to combined states include transitions of all the member microstates. Some entries are zero, indicating that no transitions were observed between the corresponding states. The statistical error in the rates is mostly around 10% (see Table VII in the Appendix).

TABLE III. Fluxes at the first interface for each metastable state obtained from the minus interface move. Columns give from left to right, the state label, the average first passage time (\overline{fpt}), its standard deviation ($stdv$), the number of paths contributed to the flux n_{path} , and the flux itself $\langle \phi_1 \rangle$ in ns^{-1} . The non-integer number of paths due to the symmetrization of paths in the ensemble.

State	aver fpt	stdv	n_{path}	$\langle \phi_1 \rangle$
N	0.5871	0.1013	4374	1.7032
PN	0.7117	0.2038	2233	1.4050
SN	0.5239	0.0297	4962	1.9089
Mg	0.3824	0.0527	2244	2.6151
meta	0.3116	0.0361	2031	3.2090
Pd	0.5423	0.1311	1953	1.8441
LN	0.5238	0.3061	854	1.9091
LSN	0.4576	0.0242	1516	2.1852
Lm	1.0480	0.2692	1487	0.9542
Lo	0.7057	0.2680	586	1.4171
I	1.1119	0.3717	1618	0.8993
W	0.9782	0.5360	1621	1.0223
Other	0.6306	0.2147	1734	1.5859
U	23.0915	1.3462	95	0.0433

TABLE IV. Full rate matrices. Rows denote leaving, columns arriving states. Top: conditional transition probability based on path type analysis, symmetrization, and WHAM. Middle: full rate matrix (in ns^{-1}) by multiply the top matrix with the fluxes. Bottom: the mean first passage time matrix (in ns), obtained from the reciprocal rates.

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	Other state	U
Conditional transition probability matrix at the first interface of each state.														
N	9.83×10^{-1}	2.18×10^{-3}	1.35×10^{-4}	2.71×10^{-4}	9.59×10^{-3}	3.10×10^{-3}	1.41×10^{-3}		6.04×10^{-5}		5.82×10^{-6}	1.23×10^{-7}	5.26×10^{-5}	1.36×10^{-5}
PN	3.95×10^{-1}	5.95×10^{-1}	3.98×10^{-4}	2.17×10^{-4}	5.10×10^{-3}	2.06×10^{-3}	1.31×10^{-3}		4.24×10^{-5}		1.20×10^{-4}		1.00×10^{-3}	2.91×10^{-5}
SN	6.02×10^{-4}	9.76×10^{-6}	9.98×10^{-1}	2.29×10^{-6}	1.47×10^{-4}	4.17×10^{-4}	1.46×10^{-5}	4.50×10^{-4}		1.30×10^{-5}	5.63×10^{-5}	1.32×10^{-8}	5.37×10^{-4}	1.15×10^{-4}
Mg	1.50×10^{-1}	6.60×10^{-4}	2.85×10^{-4}	7.05×10^{-1}	1.16×10^{-1}		2.76×10^{-2}		1.20×10^{-5}			7.93×10^{-7}	4.99×10^{-4}	
meta	2.52×10^{-1}	7.37×10^{-4}	8.68×10^{-4}	5.49×10^{-3}	7.37×10^{-1}	1.21×10^{-3}	2.58×10^{-3}	7.19×10^{-6}	1.12×10^{-4}		4.94×10^{-5}	2.82×10^{-7}	3.53×10^{-4}	2.96×10^{-5}
Pd	3.99×10^{-1}	1.46×10^{-3}	1.20×10^{-2}		5.92×10^{-3}	5.77×10^{-1}	6.90×10^{-5}	8.28×10^{-5}		1.32×10^{-4}	1.20×10^{-4}	2.10×10^{-6}	3.93×10^{-3}	6.82×10^{-5}
LN	5.58×10^{-2}	2.85×10^{-4}	1.29×10^{-4}	1.98×10^{-3}	3.90×10^{-3}	2.13×10^{-5}	9.36×10^{-1}	3.50×10^{-4}	1.19×10^{-3}		3.54×10^{-5}	4.03×10^{-6}		3.05×10^{-4}
LSN			1.48×10^{-2}		4.01×10^{-5}	9.43×10^{-5}	1.30×10^{-3}	9.81×10^{-1}		1.68×10^{-3}	4.53×10^{-5}	1.81×10^{-7}	6.44×10^{-4}	4.94×10^{-4}
Lm	8.14×10^{-2}	3.15×10^{-4}		2.91×10^{-5}	5.77×10^{-3}		4.07×10^{-2}		8.72×10^{-1}			3.65×10^{-6}		
Lo			1.72×10^{-3}			6.04×10^{-4}		6.77×10^{-3}		9.45×10^{-1}	3.06×10^{-4}	1.32×10^{-6}	3.89×10^{-2}	6.58×10^{-3}
I	1.23×10^{-2}	1.40×10^{-3}	2.68×10^{-2}		3.98×10^{-3}	1.97×10^{-3}	1.89×10^{-3}	6.54×10^{-4}		1.10×10^{-3}	9.31×10^{-1}	3.65×10^{-6}	1.22×10^{-2}	6.30×10^{-3}
W	7.91×10^{-3}		1.91×10^{-4}	9.23×10^{-5}	6.93×10^{-4}	1.05×10^{-3}	6.55×10^{-3}	7.96×10^{-5}	1.74×10^{-4}	1.44×10^{-4}	1.11×10^{-4}	8.27×10^{-1}	8.30×10^{-6}	1.56×10^{-1}
Other	6.88×10^{-3}	7.23×10^{-4}	1.58×10^{-2}	1.18×10^{-4}	1.76×10^{-3}	3.99×10^{-3}		5.75×10^{-4}		8.64×10^{-3}	7.52×10^{-4}	1.68×10^{-8}	9.59×10^{-1}	2.21×10^{-3}
U	1.94×10^{-3}	2.29×10^{-5}	3.70×10^{-3}		1.61×10^{-4}	7.58×10^{-5}	1.10×10^{-3}	4.82×10^{-4}		1.60×10^{-3}	4.25×10^{-4}	3.46×10^{-4}	2.41×10^{-3}	9.88×10^{-1}
Rate matrix (ns^{-1})														
N	...	3.71×10^{-3}	2.30×10^{-4}	4.61×10^{-4}	1.63×10^{-2}	5.28×10^{-3}	2.40×10^{-3}		1.03×10^{-4}		9.92×10^{-6}	2.09×10^{-7}	8.96×10^{-5}	2.32×10^{-5}
PN	5.55×10^{-1}	...	5.60×10^{-4}	3.04×10^{-4}	7.16×10^{-3}	2.89×10^{-3}	1.84×10^{-3}		5.96×10^{-5}		1.68×10^{-4}		1.41×10^{-3}	4.09×10^{-5}
SN	1.15×10^{-3}	1.86×10^{-5}	...	4.37×10^{-6}	2.81×10^{-4}	7.95×10^{-4}	2.78×10^{-5}	8.59×10^{-4}		2.48×10^{-5}	1.08×10^{-4}	2.52×10^{-8}	1.02×10^{-3}	2.20×10^{-4}
Mg	3.92×10^{-1}	1.73×10^{-3}	7.45×10^{-4}	...	3.03×10^{-1}		7.23×10^{-2}		3.13×10^{-5}			2.07×10^{-6}	1.30×10^{-3}	
meta	8.07×10^{-1}	2.36×10^{-3}	2.78×10^{-3}	1.76×10^{-2}	...	3.88×10^{-3}	8.28×10^{-3}	2.31×10^{-5}	3.60×10^{-4}		1.58×10^{-4}	9.07×10^{-7}	1.13×10^{-3}	9.50×10^{-5}
Pd	7.36×10^{-1}	2.69×10^{-3}	2.22×10^{-2}		1.09×10^{-2}	...	1.27×10^{-4}	1.53×10^{-4}		2.43×10^{-4}	2.21×10^{-4}	3.87×10^{-6}	7.24×10^{-3}	1.26×10^{-4}
LN	1.06×10^{-1}	5.44×10^{-4}	2.47×10^{-4}	3.78×10^{-3}	7.44×10^{-3}	4.06×10^{-5}	...	6.69×10^{-4}	2.28×10^{-3}		6.76×10^{-5}	7.70×10^{-6}		5.81×10^{-4}
LSN			3.23×10^{-2}		8.77×10^{-5}	2.06×10^{-4}	2.83×10^{-3}	...		3.68×10^{-3}	9.89×10^{-5}	3.96×10^{-7}	1.41×10^{-3}	1.08×10^{-3}
Lm	7.77×10^{-2}	3.01×10^{-4}		2.78×10^{-5}	5.51×10^{-3}		3.88×10^{-2}		...			3.48×10^{-6}		
Lo			2.44×10^{-3}			8.56×10^{-4}		9.60×10^{-3}		...	4.34×10^{-4}	1.87×10^{-6}	5.52×10^{-2}	9.33×10^{-3}
I	1.11×10^{-2}	1.26×10^{-3}	2.41×10^{-2}		3.58×10^{-3}	1.78×10^{-3}	1.70×10^{-3}	5.88×10^{-4}		9.89×10^{-4}	...	3.28×10^{-6}	1.09×10^{-2}	5.67×10^{-3}
W	8.09×10^{-3}		1.95×10^{-4}	9.43×10^{-5}	7.08×10^{-4}	1.07×10^{-3}	6.70×10^{-3}	8.14×10^{-5}	1.78×10^{-4}	1.48×10^{-4}	1.14×10^{-4}	...	8.48×10^{-6}	1.59×10^{-1}
Other	1.09×10^{-2}	1.15×10^{-3}	2.50×10^{-2}	1.87×10^{-4}	2.79×10^{-3}	6.33×10^{-3}		9.11×10^{-4}		1.37×10^{-2}	1.19×10^{-3}	2.67×10^{-8}	...	3.50×10^{-3}
U	8.42×10^{-5}	9.92×10^{-7}	1.60×10^{-4}		6.98×10^{-6}	3.28×10^{-6}	4.75×10^{-5}	2.09×10^{-5}		6.91×10^{-5}	1.84×10^{-5}	1.50×10^{-5}	1.04×10^{-4}	...

TABLE IV. (Continued.)

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	Other state	U
Mean first passage time matrix (in ns)														
N	...	269.76	4340.54	2168.61	61.24	189.23	416.91		9728.44		100 832.11	4 777 834.24	11 155.61	43 137.71
PN	1.80	...	1787.02	3285.88	139.58	345.44	544.93		16 786.75		5938.94		708.65	24 433.61
SN	870.06	53 652.22	...	228 771.80	3559.01	1257.53	36 002.95	1164.73		40 243.45	9296.81	39 707 055.43	976.15	4546.30
Mg	2.55	579.15	1343.02	...	3.30		13.83		31 985.25			482 022.16	766.38	
meta	1.24	422.86	359.12	56.80	...	257.86	120.74	43 322.70	2775.99		6314.09	1 103 103.90	883.53	10 527.97
Pd	1.36	371.55	45.05		91.55	...	7854.88	6545.74		4110.11	4516.42	258 437.55	138.11	7947.49
LN	9.39	1838.33	4045.43	264.72	134.45	24 636.16	...	1494.74	438.76		14 785.47	129 895.90		1719.91
LSN			30.92		11 398.27	4850.69	353.17	...		271.63	10 110.95	2 525 592.33	710.83	926.46
Lm	12.87	3326.07		35 956.48	181.56		25.77		...			287 000.71		
Lo			409.70			1167.97		104.16		...	2303.87	534 338.07	18.13	107.18
	90.29	796.49	41.53		279.52	563.14	587.79	1701.25		1010.89	...	304 442.69	91.40	176.52
W	123.66		5126.84	10 601.56	1411.51	931.42	149.26	12 283.06	5615.11	6776.86	8799.80	...	117 917.33	6.27
Other	91.63	871.87	40.00	5349.65	358.81	157.98		1097.20		72.96	838.44	37 424 658.23	...	285.90
U	11 880.87	1 007 937.81	6246.63		143 356.55	304 806.40	21 031.35	47 948.61		14 465.10	54 295.70	66 740.76	9586.12	...

C. Analysis of the kinetic network

1. Timescales

The reciprocal eigenvalues of the rate matrix immediately give the most important time scales in the kinetic network. These timescales are in descending order $\tau = \{\infty, 1746.44, 278.681, 25.6991, 21.3476, 16.0814, 9.96394, 8.7391, 7.56012, 5.65457, 1.74424, 1.36474, 1.27371, 1.11255\}$ ns. The slowest time scale (besides infinity, which is the equilibrium population) corresponds to the overall folding process, the second timescale is related to a relatively long lived intermediate state. All other timescales are fast processes related to transitions between the metastable states. Recent IR experiments by Meuzelaar *et al.*¹⁹ found biexponential kinetics for the Trp cage system. At a temperature of 300 K, they measured a slow time scale of 2200 ns, and a fast timescale of around 150 ns, which they attribute to the presence of an intermediate. Our prediction of these time scales are 1746 and 278 ns, respectively, in reasonable agreement with the experiments.

2. The stationary distribution or equilibrium population

The time evolution of the population $\mathbf{p}(t)$ of all 14 states can be computed from the rate matrix from $\mathbf{p}(t) = \mathbf{p}(0) \exp(\mathbf{K}t)$, where $\mathbf{p}(0)$ denotes the initial population vector. Fig. 5 shows the relaxation for a situation in which only the native state is initially populated. After full relaxation, the distribution has become stationary, the equilibrium population. For our rate matrix, the equilibrium population is given in Table V. Note that only N , U , and SN are significantly populated states, and that state SN can be considered an intermediate state in the (un)folding process. Furthermore, these populations predict that at room temperature the fraction of unfolded to folded configuration is $p_U/p_N \approx 0.23$, where we take the SN state to be a folded state. In a recent REMD calculation,⁴⁷ using the same force field, this fraction was predicted around 0.20, close to our current calculation. Previous measurements of the folding curve yielded $p_U/p_N \approx 0.33$,¹⁵ and more recent IR experiments find a $p_U/p_N \approx 0.2$, close to our predictions.¹⁹

Besides the native N and the unfolded state U , only state SN is significantly populated. The SN state is characterized by a pseudo native structure (Fig. 6) with remarkable stability, thus constitutes a major kinetic trap for correct folding. In the native state, the stable α -helix and the compact core provide the very high stability. The α -helix is stabilized by hydrogen bonds, while the hydrophobic core is stabilized by the hydrogen bond formed between W6-P18 and the D9-R16 salt bridge. In the SN state, R16, which shields the core, can rotate away due to thermal fluctuations, thus exposing W6 to solvent molecules. These subsequently penetrate the core and induce further transitions including unfolding activity.

The structural features of SN match the experimental findings by IR T-jump spectroscopy,⁵¹ where the transition state was found to have an intact α -helix, a hydrophobic core formed without the contribution from P19, and a broken D9-

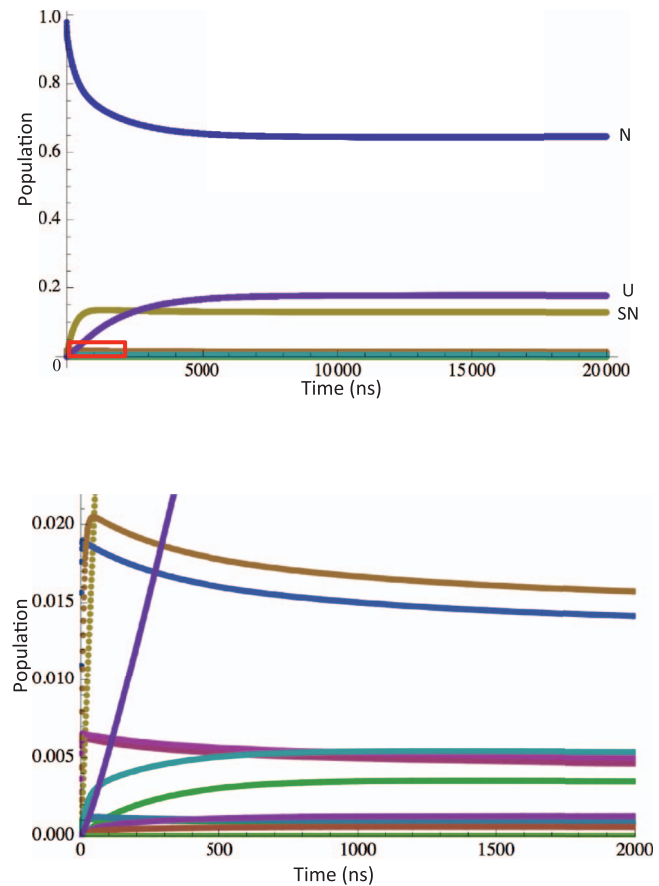


FIG. 5. Top: Plot of the population relaxation $\mathbf{p}(t) = \mathbf{p}(0)\exp(\mathbf{K}t)$, as a function of time t for all 14 states. At time $t = 0$ only the native state is populated. Bottom: Enlarged population vs. time plot of the region in red in the top figure, where the population relaxation of the less populated 11 states can be seen. Lines in the decreased order of final population represent states LN, meta, other, Pd, PN, LSN, Lo, Lm, Mg, I, and W.

R16 salt bridge. In addition, these findings corroborate with the recent IR experiment of Meuzelaar *et al.*¹⁹

Interestingly enough, we found that the intermediate I state, consisting of a formed α -helix with a solvated and

TABLE V. Equilibrium populations and committors. The left column shows the population obtained from the rate matrix, the right column the committors towards state U given by TPT analysis (see Sec. III C 4).

State	Population	Committor
N	0.64	0
PN	0.0043	0.00076
SN	0.13	0.12
Mg	0.00076	0.0014
meta	0.013	0.00094
Pd	0.0046	0.0053
LN	0.014	0.0062
LSN	0.0034	0.15
Lm	0.00085	0.0021
Lo	0.0013	0.25
I	0.00058	0.17
W	0.000017	0.90
Other	0.0053	0.16
U	0.18	1

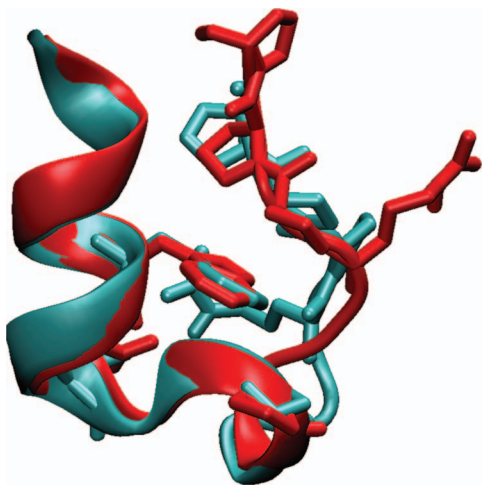


FIG. 6. Visualization of the intermediate *SN* state and the native state *N*. The *SN* state has a disordered proline tail, and a broken salt-bridge with the arginine rotated away towards the solvent.

rotated proline tail, was not significantly populated. This seems in contrast with previous findings, that proposed *I* as the key intermediate in a diffusion-collision-like unfolding channel.^{24,28} We can solve this paradox, by constructing the free energy landscape.

3. Free energy surface from the reweighted path ensemble

The free energy surface can be constructed from the path ensemble, using the concept of the reweighted path ensemble.³¹ For each state, we can reweight the paths from the MSTIS ensembles. We match the ensembles using the state populations. For simplicity, we ignore all insignificantly populated ensembles, as they will have very little effect on the free energy surface. We reweighted the path ensembles from state *N*, *SN*, and *U* and computed the free energy landscape of Trp-cage as a function of the RMSD with respect to an ideal helix (RMSD_{ahx}), and the RMSD of the α carbons ($\text{RMSD}_{C\alpha}$). The resulting landscape is shown in Fig. 7 and is comparable with that of the REMD study in Ref. 47 which uses the same force field.

From the dynamical trajectories that were used to construct this free energy landscape, it follows that most reactive unfolding paths (from *N* to *U*) first increase $\text{RMSD}_{C\alpha}$, followed by α -helix deformation and unfolding. Reversely, reactive folding paths (*U* to *N*) first fold the helix before condensing to the native state. This route passes by an intermediate ensemble of states, with $\text{RMSD}_{C\alpha} > 0.5$ nm, and $\text{RMSD}_{\text{ahx}} < 0.05$ nm, in which the proline tail points away from the hydrophobic core while the α -helix is fully formed. This intermediate region is much broader than the definition of our state *I*, which only contains a few stable structures. Transition paths can pass through the intermediate region without touching the *I* state. We identified the fraction of paths that connect the folded state to the unfolded state and found that 98% of the paths were traversing through the region with

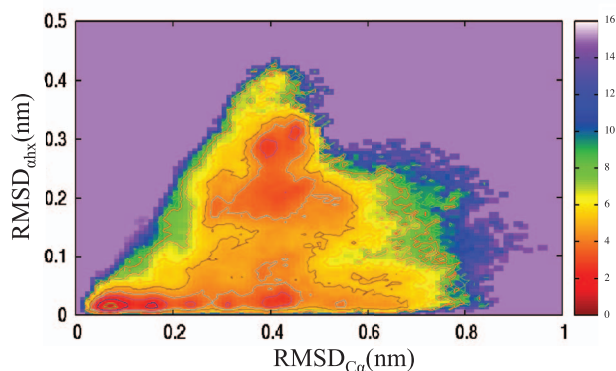


FIG. 7. Free energy landscape of Trp-cage (at 300 K) as obtained from the reweighted path ensemble and plotted on RMSD of helix against RMSD of the α carbons.

$\text{RMSD}_{C\alpha} > 0.5$ nm, and $\text{RMSD}_{\text{ahx}} < 0.05$ nm. This explains the apparent high population in this region in the free energy, and also the previous results. However, we found that structures in this region are flexible, and the trajectories do not get trapped in a single metastable structure. This also follows from the crossing probability of structure *I*, which does capture trajectories.

4. Transition path theory flux analysis

The rate matrix obtained by MSTIS contains all kinetic information of the network, but is still rather difficult to interpret. Application of concepts from TPT^{52,53} results in more insight in which transition contribute mostly to the folding or unfolding flux.³ In this flux analysis, the commitment probability, a.k.a. as committor, to reach the final state, is computed from the transition matrix. Looking at the unfolding process, we can define the initial state as *N*, the final state as state *U*. The committor q_i^+ then equals the probability that from state *i* the unfolded state *U* is reached before the native state *N*.^{6,52–54} By default $q_N^+ = 0$ and $q_U^+ = 1$. Then TPT^{3,53} gives the committors q_i^+ for all other state *i* as

$$q_i^+ = \sum_{k \in \mathbf{I}} T_{ik} q_k^+ + T_{iU}, \quad (14)$$

where T_{ik} is the transition probability to go from *i* to *k* in a certain lag time τ (set here to 1 ns), \mathbf{I} is the subset of states so that $\mathbf{M} = (\mathbf{N}, \mathbf{U}, \mathbf{I})$. Using the relation $\mathbf{T} = \exp(\mathbf{K}\tau)$ we can solve these linear equations to give the committors for the Trp-cage metastable states. The committors are given in Table V. With committor probabilities for state *i*, $j \in \mathbf{I}$, the effective flux f_{ij} from *i* to *j*, follows from³

$$f_{ij} = \pi_i q_i^- T_{ij} q_j^+, \quad (15)$$

where π_i denotes the population of state *i*, and $q_i^- = 1 - q_i^+$, the backward-committor probability, is the probability of a transition $N \rightarrow i$ rather than $U \rightarrow i$. As f_{ij} and f_{ji} are both positive, the net flux becomes³

$$f_{ij}^+ = \max\{0, f_{ij} - f_{ji}\}. \quad (16)$$

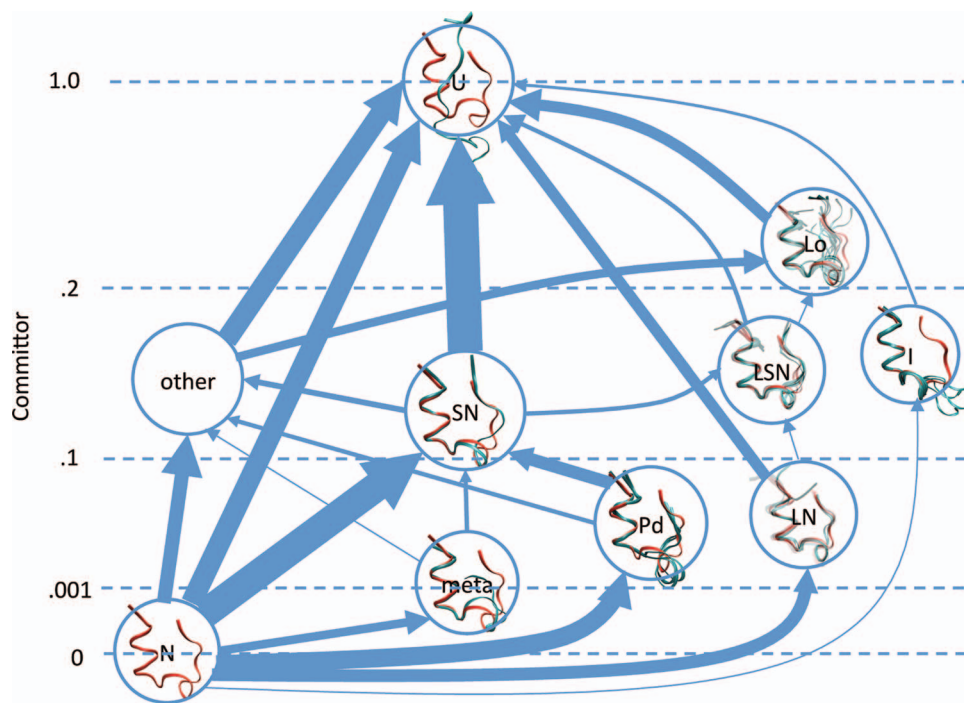


FIG. 8. Trp-cage unfolding flux. Numbers on the left indicate the committer probabilities of the corresponding levels. Thickness of the arrow indicate the $N \rightarrow U$ related flux density. This figure only regards flux greater than 1 per ms.

This net flux represents the contribution to the $N \rightarrow U$ unfolding process. The full flux matrix for the unfolding process is given in Table VIII of the Appendix. A visual representation of the unfolding flux through the kinetic network as a function of committer is shown in Fig. 8. In this plot, the thickness of the arrows between states is proportional to the net flux. This plot shows several major routes to unfolding, the direct $N \rightarrow U$, the indirect $N \rightarrow SN \rightarrow U$, and the route through the other states. State SN thus transfers a largest amount (about a third) of the net flux from state N to U , which demonstrates the key role this intermediate plays in the (un)folding process. Note that these observations also hold for the folding process. As we are describing an equilibrium kinetic network, the net folding fluxes are identical but in the opposite direction.

The TPT also gives the overall unfolding flux F (number of unfolding event per unit time, $\tau = 1\text{ns}$) and unfolding rate $k_{N \rightarrow U}$ as

$$F = \sum_{i \neq N} \pi_N T_{Ni} q_i^+, \quad (17)$$

$$k_{NU} = \frac{F}{\tau \sum_{i \in \mathbf{M}} \pi_i q_i^-}. \quad (18)$$

τ is the lag time of transition matrix, π is the population of the denoted state, and \mathbf{M} is the set containing all states. Rate k_{NU} accounts for the total unfolding rate, the sum of direct and all indirect unfolding rates. For Trp-cage unfolding, the unfolding rate constant from this analysis is k_{NU} is $(1.01 \times 10^{-4} \text{ ns}^{-1}) \approx (9.9 \mu\text{s})^{-1}$, while the overall folding rate is $4.17 \times 10^{-4} \text{ ns}^{-1} \approx (2.4 \mu\text{s})^{-1}$. These rates are in reasonable agreement with the experimental results at room temperature.

IV. CONCLUSIONS

We have resolved the kinetic network of Trp-cage in atomistic detail. Instead of a straightforward MD simulation, we used the single replica multiple state transition interface sampling, a recently developed enhanced sampling method that samples unbiased dynamical transition trajectories. The method has a great advantage of speeding up the efficiency of sampling by at least two orders of magnitude, without loss of accuracy. Given the fact that we have about a thousand (un)folding trajectories, the total time that the simulations represent is on the order of 18 ms whereas the actual spent simulation time is on the order of $150 \mu\text{s}$. The enhancement comes from the fact that a protein kinetic network has many basins of attraction that temporally trap the system, whereas the MSTIS biases these trapped trajectories to regions where they are free to explore the rest of the configurational space.

The method results in a full rate matrix for all metastable states. For the Trp-cage, this meant a 14×14 rate matrix. The results indicate the presence of a metastable near native intermediate state SN in the Trp-cage network, characterized by a full alpha helix, but a slightly disordered proline tail, a broken salt bridge, and a rotated arginine residue. This SN is an important transient state in the network, with a population of about 13%. Our prediction shows it carries a third of the net flux through the network. The overall rates are in good agreement with previous simulations, and with experiment. Moreover, the presence of the intermediate was recently found in extensive T-jump IR experiments.¹⁹ Also, the predicted timescales are in agreement with these experiments.

Remarkably, while most of the flux is carried through a few routes in the network, the rate matrix is not sparse. While

some transitions were not observed, the majority of the entries in the rate matrix are filled, meaning that transitions exist. This is in contrast to the standard MSM approaches⁴ which usually result in very sparse transition matrices. This difference is naturally caused by our rather strict metastable state definitions, in combination with the fact that we allow relatively long paths between states. Therefore, paths are not required to visit neighboring states, but also can connect conformations that are further apart.

Besides having similarities to the MSM methodology, the MSTIS methodology at first sight also resembles the cut-based procedure of Krivov and Karplus to obtain equilibrium kinetic networks for proteins.^{55,56} However, the methods are different in spirit and in approach. The cut-based method has the advantage that it needs no prior information of the stable states, whereas MSTIS requires knowledge of the stable states. An advantage of MSTIS is that it can handle large barriers and long timescales, whereas the cut-based methodology is based on an equilibrium MD trajectory, which might be computationally not tractable when long timescales are involved.

As a final remark we note that our work shows the feasibility of studying the full kinetic network of a (small) model protein in atomistic detail using modest computational resources, and thus opens the way for studying larger, biologically more relevant systems.

ACKNOWLEDGMENTS

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

APPENDIX: CONVERGENCE OF THE BIAS FUNCTIONS AND ADDITIONAL KINETIC DATA

Fig. 9 shows the last three fixed bias functions ($\ln P(\lambda)$) for the states, N, PN, SN, Mg, meta, Pd, I, W, and U. Also shown is the final crossing probability. Overlap of the final curves imply convergence. Tables VI–VIII, show the symmetrized count matrix, the relative standard deviation of the rate matrix, and the flux matrix.

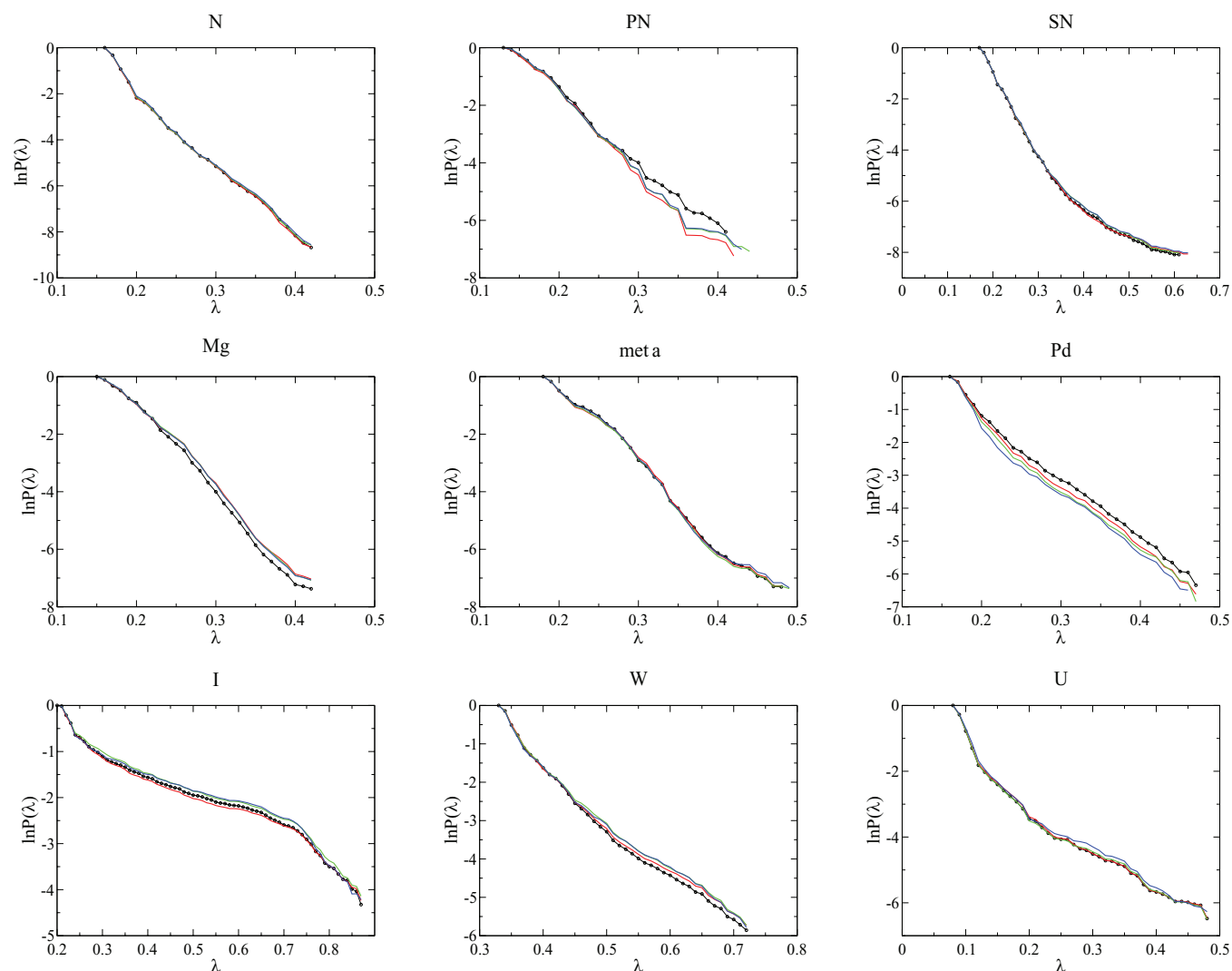


FIG. 9. The last three fixed biasing functions (in the order of blue, green and red, red being the latest bias condition) and the final crossing probability (black dots) for each of the sampled states.

TABLE VI. Symmetric path count matrix (top table) and average path length matrix (bottom table) calculated based on all sampled paths.

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	Other	U
Path number matrix gives the number of paths for the transition of the corresponding two states														
N	145 88	4923	464	4240	9132	2325	1644	1	768	32	276	471	286	200
PN		2018	39	29	322	68	120	0	5	0	28	0	125	3
SN			16 218	61	355	802	13	214	6	61	507	18	1373	348
Mg				3080	1236	0	529	0	6	0	0	4	40	0
meta					2666	261	109	2	106	11	85	53	186	8
Pd						2426	7	8	0	28	63	79	369	11
LN							2504	30	188	13	49	413	15	68
LSN								2142	5	17	18	8	15	39
Lm									1434	20	1	18	29	0
Lo										626	22	7	280	85
I											5850	10	397	177
W												4156	1	1651
other													5770	170
U														3702
Path length (in ns) matrix gives the average path length for a corresponding transition														
N	1.10	1.59	1.92	1.02	0.55	5.87	2.25	19.41	1.81	2.91	4.86	8.19	4.05	17.63
PN		0.70	0.91	2.28	1.24	2.62	4.96	0.00	4.22	0.00	10.38	0.00	2.26	15.63
SN			1.04	0.54	0.89	1.00	17.43	2.87	5.83	6.46	3.40	9.41	2.46	10.90
Mg				0.60	0.87	0.00	1.59	0.00	5.13	0.00	0.00	4.43	1.31	0.00
meta					0.67	0.85	2.14	5.14	1.90	2.33	6.06	7.09	1.87	31.71
Pd						1.05	6.69	12.27	0.00	1.64	7.15	4.36	2.80	7.10
LN							2.24	3.76	3.66	3.70	13.86	9.84	17.16	20.13
LSN								0.46	6.59	5.35	12.70	1.86	7.65	11.12
Lm									0.78	10.02	37.53	14.85	6.95	0.00
Lo										0.42	14.68	14.00	1.97	11.50
I											3.14	27.05	5.58	22.80
W												3.21	11.05	5.81
Other													1.92	17.37
U														4.60

TABLE VII. Relative standard deviation matrix for the rate matrix from 10 sets of data. For each transition at the first interface, a virtual transition path number was generated from a Gaussian distribution around the average path number, with a width of the square root of the path number.

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	Other	U
N	...	0.1006	0.0894	0.0151	0.0117	0.0387	0.0406		0.0426		0.0772	0.1066	0.1381	0.1122
PN	0.0025	...	0.0669	0.0738	0.0260	0.0551	0.0442		0.3088		0.1266		0.3539	0.2617
SN	0.0646	0.1571	...	0.1289	0.0673	0.0514	0.4852	0.1041		0.0848	0.0447	0.1546	0.0694	0.0580
Mg	0.0036	0.0617	0.1014	...	0.0053		0.0096		0.3759			0.1852	0.0670	
meta	0.0029	0.1024	0.0284	0.0102	...	0.0478	0.0286	0.7030	0.0300		0.0569	0.1052	0.0998	0.1474
Pd	0.0015	0.0928	0.0106		0.0127	...	0.2148	0.1981		0.1174	0.0567	0.1107	0.0655	0.1328
LN	0.0064	0.0677	0.4938	0.0295	0.0598	0.2922	...	0.0780	0.0257		0.0711	0.0999		0.1176
LSN			0.1108		1.9173	0.2558	0.0489	...		0.0916	0.0698	0.1810	0.1335	0.0904
Lm	0.0057	0.2581		1.2667	0.0248		0.0078		...			0.1231		
Lo			0.0392			0.1232		0.0246		...	0.0855	0.1736	0.0065	0.0871
I	0.0140	0.0521	0.0087		0.0224	0.0292	0.0313	0.0548		0.0490	...	0.2028	0.0157	0.0203
W	0.0151		0.1342	0.1997	0.0704	0.0490	0.0225	0.2243	0.0985	0.1113	0.1649	...	2.5636	0.0041
Other	0.0207	0.3776	0.0140	0.1361	0.0325	0.0635		0.0735		0.0142	0.0686	0.8700	...	0.0877
U	0.0378	0.2608	0.0335		0.1128	0.1135	0.0571	0.0994		0.0444	0.0692	0.0201	0.0355	...

TABLE VIII. Flux matrix of 14 states for transition from N to U in units per ms.

	N	PN	SN	Mg	meta	Pd	LN	LSN	Lm	Lo	I	W	Other	U
N	1.37	22.46	0.36	6.63	12.34	9.12	0.13	0.13	0.20	1.26	0.12	11.14	15.96	
PN		0.27		0.01	0.05	0.04			0.01	0.09		0.72	0.17	
SN							3.00		0.56	0.72		4.93	25.29	
Mg		0.08				0.18						0.12	0.02	
meta		3.09	0.04		0.17	0.43	0.04		0.02	0.24	0.01	1.61	0.99	
Pd		8.17					0.08		0.23	0.12	0.01	3.45	0.50	
LN		0.44					1.26			0.15	0.09	0.03	7.93	
LSN									1.28	0.01		0.05	3.16	
Lm						0.12							0.01	
Lo													8.90	
I									0.05				2.64	
W													0.24	
Other									6.55	0.09			15.41	
U														

- ¹D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, *Science* **330**, 341 (2010).
- ²V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, *J. Am. Chem. Soc.* **132**, 1526 (2010).
- ³F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weigl, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19011 (2009).
- ⁴J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *J. Chem. Phys.* **134**, 174105 (2011).
- ⁵C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998).
- ⁶P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- ⁷P. G. Bolhuis and C. Dellago, *Rev. Comput. Chem.* **27**, 102 (2009).
- ⁸T. S. van Erp, D. Moroni, and P. G. Bolhuis, *J. Chem. Phys.* **118**, 7762 (2003).
- ⁹J. Rogal and P. G. Bolhuis, *J. Chem. Phys.* **129**, 224107 (2008).
- ¹⁰W.-N. Du, K. A. Marino, and P. G. Bolhuis, *J. Chem. Phys.* **135**, 145102 (2011).
- ¹¹T. S. van Erp, *Phys. Rev. Lett.* **98**, 268301 (2007).
- ¹²P. G. Bolhuis, *J. Chem. Phys.* **129**, 114108 (2008).
- ¹³W. Du and P. G. Bolhuis, *J. Chem. Phys.* **139**, 044105 (2013).
- ¹⁴J. Neidigh, R. Fesinmeyer, and N. Andersen, *Nat. Struct. Biol.* **9**, 425 (2002).
- ¹⁵L. Qiu, S. A. Pabit, A. Roitberg, and S. Hagen, *J. Am. Chem. Soc.* **124**, 12592 (2002).
- ¹⁶W. W. Streicher and G. I. Makhatadze, *Biochemistry* **46**, 2876 (2007).
- ¹⁷H. Neuweiler, S. Doose, and M. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16650 (2005).
- ¹⁸Z. Ahmed, A. M. I. A. Beta, and S. A. Asher, *J. Am. Chem. Soc.* **127**, 10943–10950 (2005).
- ¹⁹H. Meuzelaar, K. A. Marino, A. Huerta-Viga, M. R. Panman, L. J. Smeenk, A. J. Kettlerij, J. H. van Maarseveen, P. Timmerman, P. G. Bolhuis, and S. Woutersen, *J. Phys. Chem. B* **117**, 11490 (2013).
- ²⁰C. D. Snow, B. Zagrovic, and V. S. Pande, *J. Am. Chem. Soc.* **124**, 14548 (2002).
- ²¹R. Zhou, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13280 (2003).
- ²²S. Chowdhury, M. C. Lee, and Y. Duan, *J. Phys. Chem. B* **108**, 13855 (2004).
- ²³J. Juraszek and P. Bolhuis, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15859 (2006).
- ²⁴F. Marinelli, F. Pietrucci, A. Laio, and S. Piana, *PLoS Computational Biology* **5**(8), e1000452 (2009).
- ²⁵J. Juraszek and P. Bolhuis, *Biophys. J.* **95**, 4246 (2008).
- ²⁶C. Velez-Vega, E. E. Borrero, and F. A. Escobedo, *J. Chem. Phys.* **133**, 105103 (2010).
- ²⁷K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, *Science* **334**, 517 (2011).
- ²⁸J. Juraszek, G. Saladino, T. S. van Erp, and F. L. Gervasio, *Phys. Rev. Lett.* **110**, 108106 (2013).
- ²⁹R. Day, D. Paschek, and A. E. Garcia, *Proteins* **78**, 1889 (2010).
- ³⁰P. G. Bolhuis, *J. Phys. Condens. Matter* **15**, S113 (2003).
- ³¹J. Rogal, W. Lechner, J. Juraszek, B. Ensing, and P. G. Bolhuis, *J. Chem. Phys.* **133**, 174109 (2010).
- ³²S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).
- ³³F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- ³⁴F. Wang and D. P. Landau, *Phys. Rev. E* **64**, 056101 (2001).
- ³⁵B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).
- ³⁶D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *J. Comput. Chem.* **26**, 1701 (2005).
- ³⁷E. Lindahl, B. Hess, and D. van der Spoel, *J. Mol. Mod.* **7**, 306 (2001).
- ³⁸H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, *Comput. Phys. Commun.* **91**, 43 (1995).
- ³⁹V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Proteins* **65**, 712 (2006).
- ⁴⁰W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- ⁴¹G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- ⁴²B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, *J. Comput. Chem.* **18**, 1463 (1997).
- ⁴³T. Darden, D. York, and L. G. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).
- ⁴⁴U. Essmann, L. Perela, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- ⁴⁵Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- ⁴⁶D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed., Computational Science Series (Academic Press, San Diego, 2002).
- ⁴⁷K. A. Marino and P. G. Bolhuis, *J. Phys. Chem. B* **116**, 11872 (2012).
- ⁴⁸R. A. Jarvis and E. A. Patrick, *IEEE Trans. Comput. C-22*, 1025 (1973).
- ⁴⁹T. Gonzalez, *Theor. Comput. Sci.* **38**, 293 (1985).
- ⁵⁰S. Dasgupta and P. M. Long, *J. Comput. Syst. Sci.* **70**, 555 (2005).
- ⁵¹R. Culik, A. Serrano, M. Bunagan, and F. Gai, *Angew. Chem., Int. Ed.* **50**, 10884 (2011).
- ⁵²Weinan E and E. Vanden-Eijnden, *J. Stat. Phys.* **123**, 503 (2006).
- ⁵³W. E and E. Vanden-Eijnden, *Annu. Rev. Phys. Chem.* **61**, 391 (2010).
- ⁵⁴R. Du, V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich, *J. Chem. Phys.* **108**, 334 (1998).
- ⁵⁵S. V. Krivov and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14766 (2004).
- ⁵⁶S. V. Krivov, S. Muff, A. Cafilisch, and M. Karplus, *J. Phys. Chem. B* **112**, 8701 (2008).