JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

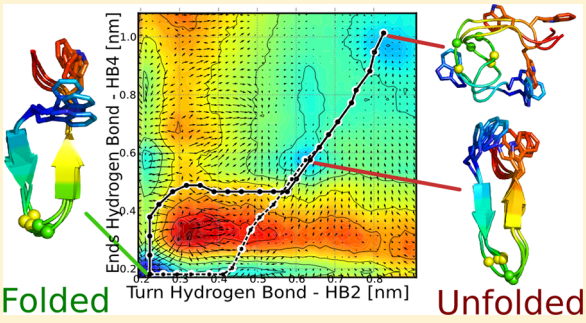
# Reconstructing the Most Probable Folding Transition Path from Replica Exchange Molecular Dynamics Simulations

Camilo Andres Jimenez-Cruz and Angel E. Garcia*

Department of Physics, Applied Physics and Astronomy, and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

Supporting Information

**ABSTRACT:** The characterization of transition pathways between long-lived states, and the identification of the corresponding transition state ensembles are useful tools in the study of rare events such as protein folding. In this work we demonstrate how the most probable transition path between metastable states can be recovered from replica exchange molecular dynamic simulation data by using the dynamic string method. The local drift vector in collective variables is determined via short continuous trajectories between replica exchanges at a given temperature, and points along the string are updated based on this drift vector to produce reaction pathways between the folded and unfolded state. The method is applied to a designed beta hairpin-forming peptide to obtain information on the folding mechanism and transition state using different sets of collective variables at various temperatures. Two main folding pathways differing in the order of events are found and discussed, and the relative free energy differences for each path estimated. Finally, the structures near the transition state are found and described.




## 1. INTRODUCTION

A protein adopts its functional shape through a series of complex solvent-influenced events and conformational changes.[1] This mechanism, known as protein folding, is now commonly understood as a slow process occurring via rare events in a high-dimensional configurational space with multiple metastable states. All-atom, explicit solvent simulations of these processes can provide detailed information about their dynamic nature and hence have proven to be a valuable tool in the study of protein folding.[2−4]

However, current computational power and methods make it expensive for a conventional molecular dynamics (MD) simulation to obtain enough statistics on relevant time scales, even for relatively small proteins. While the fastest degrees of freedom are in the order of femtoseconds, many interesting biomolecular processes involve time scales of microseconds to seconds. This discrepancy has motivated the development of several strategies to generate and analyze simulated data. For a small, single-domain protein with a funneled energy landscape, a significant fraction of reactive trajectories is expected to be distributed among narrow tubes connecting the folded and unfolded states.[5,6] By choosing appropriate collective variables on which to project the free energy landscape, the sampling can be focused on relevant portions of the configurational space to obtain information about these reactive trajectories. These relevant regions can be found using methods such as nudged elastic band,[7] transition path sampling,[8] dynamic importance sampling,[9] umbrella sampling,[10] steered molecular dynamics,[11] and string methods.[12]

Other approaches aim to improve the sampling of the configurational space such as replica exchange molecular dynamics (REMD),[13,14] which enhances the efficiency of sampling accurate canonical populations by using several replicas of the system, arranged in a temperature ladder. Periodically, exchanges are attempted between neighboring replicas and are accepted with a Metropolis probability. REMD allows replicas to perform a random walk in temperature space, enhancing the sampling of the free energy landscape at low temperatures by using high temperatures to explore new regions of the configurational space. While this greatly improves equilibrium sampling for each replica, the exchanges result in discontinuous trajectories at a given temperature, discouraging the use of REMD data to extract useful kinetic information.

However, some strategies have been proposed to obtain kinetic information from REMD data.[5,15−19] Here we show how REMD data can be used along with the dynamic string method[6,20] in order to obtain the most probable transition path at a given temperature without making assumptions on the long-term behavior of the system[6] as in the case of jump processes such as Markov Models. Our approach allows us to use a moderate number of collective variables to describe reactive pathways on the potential of mean force at the sampled temperature, with no extra cost on the simulation time, given enough sampling and a reasonable choice of parameters.

**Received:** March 5, 2013
**Published:** June 26, 2013

   dx.doi.org/10.1021/ct400170x | J. Chem. Theory Comput. 2013, 9, 3750−3755

To illustrate its use, the method is applied to folding/unfolding equilibrium REMD data of a 14 amino acids long peptide that folds into a stable beta hairpin. The simulated peptide contains two tryptophans arranged in a face to edge contact in the native state and was found to be remarkably stable in water.[22] Different geometrical measures are used as collective variables, showing the existence of two main paths between the unfolded and folded states differing in the order of events. Both paths undergo similar collapse and contain residual secondary structure in the unfolded state. The free energy differences along the paths are calculated. At the same time, structures representing the free energy minima and energy barriers are discussed.

## 2. METHODS

In REMD,[13,14] several copies of the system are simulated at different temperatures. Periodically, exchanges of temperature between neighboring replicas are attempted and accepted with probability

$$P(i \leftrightarrow j) = \min(1, \exp[(\beta_j - \beta_i)(U_j - U_i)]) \quad (1)$$

where $\beta_i = 1/k_B T_i$ is the inverse temperature and $U_i$ is the potential energy of the $i$th replica. In this way, each replica performs a random walk in the temperature space which accounts for discontinuous trajectories at a given temperature. Moreover, in order to maximize the efficiency of the REMD approach, shorter exchange times $\tau_{eX}$ are preferred resulting in $\tau_{eX}$ much shorter than the relevant time scales for the biomolecules. However, in between exchanges, the trajectories at a given temperature are formally exact.

Following the work fo Buchete and Hummer,[15] we use the short continuous trajectories from REMD simulations to update a string based on the drift velocity as proposed by Pan et al.[20] The goal of the string method, originally proposed by E et al,[12] is to obtain a sequence of configurations connecting the reactant and the product in a particular reaction. Different update rules have been proposed and lead to different paths such as the minimal energy path,[12] minimum free energy paths,[21] and most probable transition path.[20]

In the work presented here, the string corresponds to a smooth curve $\psi$ in the space spanned by a set of collective variables $\vec{z}_i(\vec{x}_i)$ constructed using a priori insight. The curve is parametrized by $\alpha$ with $\alpha \in [0, 1]$ and is initialized such that $\alpha = 0$ is the product and $\alpha = 1$ is the reactant.

The string is updated by discretizing it into $N$ image points whose dynamics in the feature space are assumed to be correctly described by a diffusion process. For an infinitesimal time step $\delta t$, the position $\vec{z}_i$ of the $i$th point follows the stochastic equation[6]

$$\Delta\vec{z}_i(t) = \vec{z}_i(t + \delta t) - \vec{z}_i(t) \quad (2)$$

$$= (-\beta \mathbf{D}\nabla V + (\nabla\cdot\mathbf{D})^T)\delta t + \sqrt{2\delta t}\,\sigma\vec{R} \quad (3)$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature, $\mathbf{D}$ is the position-dependent diffusion tensor, $V$ is the potential, $\sigma$ is related to $\mathbf{D}$ by $\mathbf{D} = \sigma\sigma^T$ and $\vec{R}$ is a vector of uncorrelated Gaussian random variable with zero mean and unit standard deviation. In eq 3, the first term of the right-hand side is deterministic and known as drift. The term multiplying the white noise is the volatility and accounts for the stochasticity of the equation. Periodically, the string is reparameterized to ensure desired constraints like smoothness and equal distance between image points, avoiding the lumping of image points near to the basins.

A careful characterization of the dynamic string method was recently done by Johnson and Hummer,[6] which showed how the most probable transition path is less sensitive to the particular choice of collective variables than other variants of the string method and generates pathways that follow the reactive flux. The dynamic string method evaluates the local drift vector at each image point by averaging the displacement in collective variables over an ensemble of unbiased short trajectories that start from the desired point. This is equivalent to taking the ensemble average of eq 3 and constraining the collective variables $\vec{z}_i(t)$ to the coordinates of the image point. The estimate of the local drift vector averages out the random fluctuations caused by thermal noise without making any assumptions about the dynamics in the underlying high dimensional space. The update procedure is repeated until the string relaxes into a stable path which is denoted as the most probable transition path (MPTP), corresponding to the zero-temperature path on the potential of equilibrium mean force for the collective variables at the sampled temperatures. Since the short trajectories between replica exchanges obey regular MD at a given temperature, they can be used to evaluate the local drift vector, and thus, dynamic string method can be applied as a posterior analysis of equilibrium REMD data.

The proposed method discretizes the feature space in cells by simple binning. The short trajectories from the REMD data are then projected onto the feature space and each frame is assigned to a corresponding cell. Then, for each image point of the string in the feature space, a random sample of frames from the REMD data occupying the same cell is drawn. The local drift vector is estimated by averaging the next frame in the trajectory and the image point moved in this direction.

In the present work, each collective variable was divided into 100 bins. For each cell, the average drift vector was obtained by finding the average displacement after 1 ps. For each image of the string, the drift vector was interpolated using neighboring cells. More details on the implementation can be found in the Supporting Information.

The free energy of a point $\vec{z}_0$ in the feature space can be calculated by finding the potential of equilibrium mean force $W(\vec{z}_0)$ for the collective variables

$$\exp(-\beta W(\vec{z}_0)) = \frac{\int d\vec{X}\,\delta(\vec{z} - \vec{z}_0)\exp(-\beta U(\vec{X}))}{\int d\vec{X}\,\exp(-\beta U(\vec{X}))} \quad (4)$$

Assuming uniform density of states in the (discrete) feature space, the free energy of a point $\alpha$ along the string reduces to

$$G(\alpha) = G(\vec{\psi}(\alpha)) \sim -k_B T \log[\pi(\psi(\alpha))] + G_0 \quad (5)$$

where $\pi(\psi(\alpha))$ is the equilibrium probability of finding the system in the cell corresponding to $\psi(\alpha)$ and $G_0$ is a constant. This probability can be easily computed from equilibrium REMD data.

Another quantity of interest is the committor function $q(\vec{z})$. The committor is defined as the probability that the system, being at the point $\vec{z}$ in the collective variables, reaches the product state before reaching the reactant state. The points $\vec{z}$ where the committor is $q(\vec{z}) = 0.5$ have equal probability of ending in either state and thus are identified as the transition state ensemble (TSE).

Assuming that the reactive flux and the local drift vector are parallel, and that the reactive flux is constant along the path, an approximation for the committor along the path can be found:[6]

$$q(s) \sim \frac{\int_0^s \mathrm{d}s' D^{-1}(s') \exp(\beta W(s'))}{\int_0^1 \mathrm{d}s' D^{-1}(s') \exp(\beta W(s'))} \quad (6)$$

where $D^{-1}(s) = \vec{t}^{\mathrm{T}} \vec{D}^{-1} \vec{t}$, $\vec{t}$ being a unit vector tangent to the path. Here, the drift tensor was taken to be proportional to the covariance matrix of the end points of all the drift trajectories.

The peptide with sequence Ace-WVSINGKKIWTG-NH2 was used as a nontrivial model system to test the method described above. This peptide adopts a stable beta hairpin configuration with a tight turn and four interstrand backbone hydrogen bonds.[22] The ends of the beta hairpin are stabilized by the face to edge contact between the two tryptophans and hydrogen bonds between the acetyl in the N terminal and the threonine and glycine in the C terminal, as shown in Figure 1.

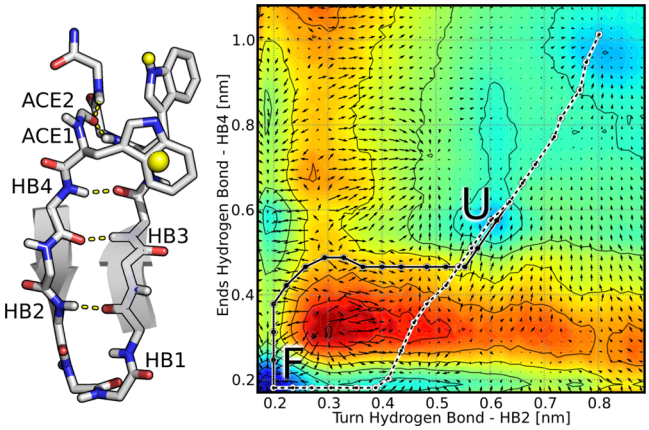



**Figure 1.** Native structure (left). The beta hairpin structure contains four interstrand backbone hydrogen bonds and a tight turn. The end is stabilized by two tryptophans arranged in a face to edge configuration, with the N-terminal tryptophan as the face. The acetyl forms hydrogen bonds with the threonine and glycine in the C-terminal. Collective variables used to describe the structure in this study are the distance between the atoms involved in HB2, HB4, ACE2, and the distance between the yellow spheres. Main reactive paths (right). An FEL (free energy landscape) in the collective variables is shown, with blue denoting lower free energy. The method is sensitive to initial conditions and can yield different results depending on the starting string.[21] Both strings follow similar collapses between basins A and B. The main difference is the order of events. For the solid string, the hydrogen bond closest to the turn forms first and the ends of the beta hairpin follow. For the dashed string, this order is reversed.

Simulation details, along with thermodynamic equilibrium characterization can be found in the Appendix.

## 3. RESULTS

The dynamic string method was used to generate the most probable transition path from REMD data by monitoring different observables as collective variables at $T$ = 334 K. First, two interstrand hydrogen bonds (HB2 and HB4 in Figure 1) were chosen because they describe to a great extent the formation of the turn and the closing of the ends of the beta hairpin at a backbone level. The free energy landscape projected on those two observables can be seen in Figure 1. The basin in the lower left corner represents configurations where both hydrogen bonds are formed, and therefore it is considered the folded state. Several runs of the method with different initial conditions converged to two different paths shown. Here, both paths undergo similar collapse from extended structures into the two basins located at (HB2, HB4) = (0.8, 1.0) and (0.6, 0.55) nm. After this collapse, two different folding scenarios are observed differing in the order of events: the solid path shows turn formation before $\beta$ strand organization, while the dashed path forms the hydrogen bond corresponding to the hydrophobic core first and then stabilizes the turn. It is also clear that the most probable transition path for the dashed string does not correspond to the minimum free energy path which would go through the saddle point located around (HB2, HB4) = (0.8, 0.3) nm. This is a result of the local diffusion coefficient being anisotropic, and thus, the string sacrifices free energy for speed and therefore climbs uphill at some points.

Representative structures for points along the string showed large differences in packing of the hydrophobic core and position of the termini, suggesting the need of using other collective variables in order to obtain a better characterization of the MPTP. Accordingly, the distance between the $\varepsilon$ hydrogen of the C-terminal tryptophan and the center of mass of the indole ring of the N-terminal tryptophan, yellow spheres in Figure 1 (left), along with the distance between the acetyl and the C-terminal glycine were used as collective variables in addition to the two hydrogen bonds described previously. Since adding more dimensions to the feature space reduces the volumetric density of points, five temperatures ranging from 324 to 344 K were employed simultaneously to extract the short trajectories, improving in this way the statistics. The resulting strings from using different initial conditions are shown in Figure 2.

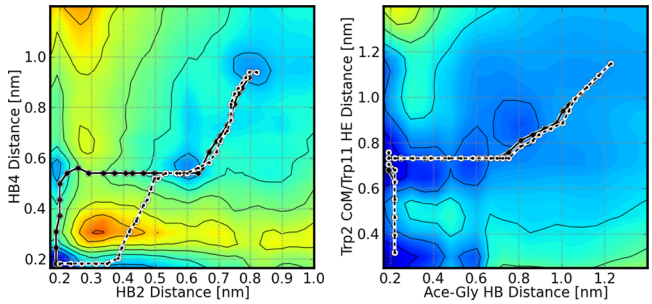



**Figure 2.** Strings found using four collective variables. The distance between the atoms involved in the hydrogen bonds HB2, HB4, and ACE2, as well as the distance between the tryptophans, Figure 1, were used to construct the feature space. The 4D paths are projected into two pairs of variables for visual inspection. The use of the new collective variables allows the string to distinguish between very different configurations which leads to a more accurate description of the MPTP and better characterization of the structures along the path.

The free energy along the new paths was computed as described in the Methods section and is shown in Figure 3. It was found that the estimate of the free energy barrier is sensitive to having enough image points close to the high energy region, which can be easily fixed in the reparametrization step by using more images. It was also found that the estimate varies considerably when using different radius for the calculation of the populations at each image point, Supporting Information Figure S4. Thus, this method for calculating free

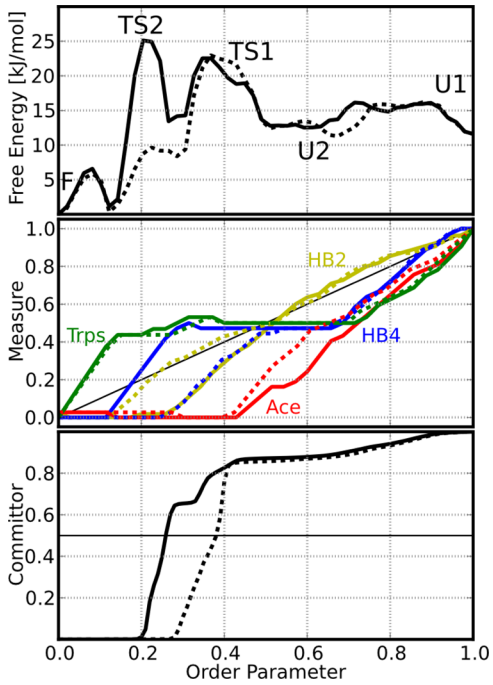



**Figure 3.** (top) Free energy profiles. Both paths contain several local minima which correspond to steps in the folding process. The path where the strands align first (dashed path) has the lower activation energy ($E_A \sim 9$ kJ/mol) than the path with early forming of the hydrogen bond at the turn ($E_A \sim 13$ kJ/mol). (middle) Projection of the collective variables (solid: turn first). The acetyl−glycine hydrogen bond closes first followed by the strand alignment and final reorganization of the tryptophans. The main difference between the two paths corresponds to the order of formation of the backbone hydrogen bonds HB2 and HB4. (bottom) Committor probability. The probability of reaching the unfolded state before reaching the folded state was calculated along the path. The points with committor values of 0.5 are identified as the transition state and correspond with the free energy peaks.

energy provides only an approximate measure for the activation energy, but allows for comparison of ratios between paths. For this peptide, the dashed transition pathway shows a free energy barrier of 9 kJ/mol from the unfolded state U2, while the solid path has a higher activation energy (15 kJ/mol).

A normalized projection of each collective variable along the string for the paths can be found in Figure 3 and provides good insight on the folding mechanism. The initial collapse from U1 to U2 consists of closing the ends of the structure by both approaching the tryptophans and joining the termini together. After the initial collapse, the acetyl−glycine hydrogen bond continues to form and reaches its final distance first in the folding process (order parameter = 0.4). From U2, the dashed path continues closing HB4 and then HB2, while for the solid path this is reversed. The last step consists on the rearrangement of the tryptophans into the native configuration reaching finally the native structure. Finally, the committor was calculated and is shown in the same figure. The committor change occurs rapidly along the order parameter for both paths, and the transition states (identified by committor values of 0.5) correspond with peaks of free energy. Among the measures used, HB4 has strong correlation with the committor for values between 0.0 and 0.65, after which HB2 becomes more relevant.

Finally, representative structures of the points labeled in Figure 3 were drawn from the ensemble sampled by the REMD simulation and are shown in Figure 4. For visual reference, the $C_\alpha$ carbon of the asparagine and glycine in the turn are shown

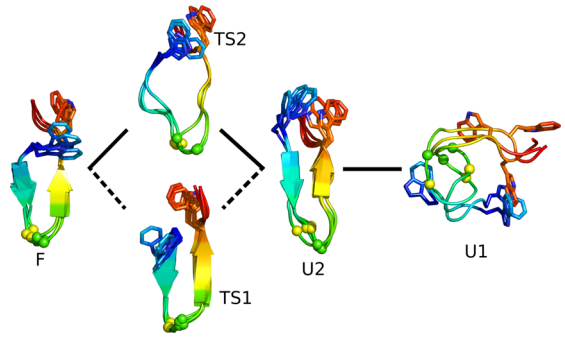



**Figure 4.** Representative structures along the strings. The unfolded state of both strings contains the minima U1 and U2, with U1 showing long turn regions. The solid path corresponds to the formation of the turn hydrogen bond before the arrangement of the hydrophobic core native contacts. The peak of free energy for both paths are shown (TS1 and TS2).

in yellow and green, respectively. The unfolded ensemble contains two main metastable states labeled U1 and U2. U1 contains long turn regions and the tryptophans in opposite sides completely exposed to the solvent. Structures in U2 have some interstrand hydrogen bonds resembling hairpins, but neither the turn or the end contacts are formed yet. Moving from U2, two paths are possible: in the dashed path, the hydrogen bonds near the middle of the hairpin are formed but the C-terminal tryptophan needs to flip in order to reach folded state. For the solid path, the state TS2 already has the tryptophans resembling the native state but needs to organize the beta strands before folding.

## 4. CONCLUDING REMARKS

Biomolecules undergo complex conformational changes relevant for their biological activity and function. Reactive trajectories between these conformations provide insight into the mechanism and rates of the underlying processes. Here, we have shown how to obtain such trajectories by using equilibrium REMD data in conjunction with the dynamic string method. Using two and four collective variables for a short beta hairpin in explicit solvent, the method allowed for the generation of the two most probable transition paths between the folded and unfolded states, as well as the identification and characterization of the conformations at the high energy peaks of the free energy profiles.

For the peptide studied here, the folding process starts with a collapsed unfolded structure exposing the tryptophans to the solvent (U1). Closing of the ends and approaching of the tryptophans follows, generating a beta hairpin structure with the turn shifted one aminoacid from the native state (U2). Alignment of the strands toward the folded structure starts from either end of the hairpin (TS1 and TS2) and finishes with the reorganization of the tryptophans to produce the correct packing (F).

Given enough sampling, the proposed method can be applied efficiently with a moderate number of collective variables at no extra simulation cost. This becomes increasingly relevant when dealing with more complex systems where one or two collective variables cannot be devised a priori, and the visual inspection of

free energy landscapes on several dimensions becomes more challenging.

Since no extra simulation cost is required, this method allows for inexpensive evaluation of collective variables as reaction coordinates. It can also be used iteratively in conjunction with statistical techniques such as principal component analysis[28] to identify and refine collective variables. The dependence on temperature of the diffusion tensor is also of interest. The method was applied to different temperatures, and a change in the preferred mechanism for strings starting from the basin U2 and ending in the folded state was observed (data not shown). Future work can explore this temperature dependence quantitatively.

Finally, the enhancement in sampling provided by REMD has been shown to increase with decreasing exchange times.[29] Since the accuracy on the calculation of the position dependent drift vector depends on the availability of continuous trajectories at each point, shorter exchange times reduce the usable information to update the strings and thus an appropriate exchange is required in order to use this method.

## APPENDIX

Temperature replica exchange molecular dynamics (REMD) as implemented in Gromacs[23] was used to sample the equilibrium ensemble of configurations of the peptide Ace-WVSINGK-KIWTG-NH2 solvated in a cubic box of 4.03 nm side with 2074 water molecules and two $Cl^-$ ions to neutralize the total charge of the system. Coulomb interactions were approximated by using PME with a cubic 54 × 54 × 54 grid and truncated at 0.9 nm. The Amber99SB[24] force field was used to model the protein atoms, and TIP3P, to model the water molecules.[25]

The REMD simulation was started from an extended structure built with Pymol,[26] solvated and equilibrated during 200 ps in the *NPT* ensemble at pressure $P$ = 1 atm and $T$ = 300 K. The last frame of the *NVT* equilibration was used as initial structure for the REMD simulation. A total of 40 replicas with temperatures 275, 279.0, 283.1, 287.3, 291.6, 295.9, 300.4, 304.9, 309.5, 314.2, 319.0, 323.9, 328.9, 334.0, 339.2, 344.5, 350.0, 355.5, 361.2, 367.0, 372.9, 379.0, 385.1, 391.4, 397.9, 404.5, 411.2, 418.1, 425.1, 432.2, 439.5, 447.0, 454.6, 462.3, 470.2, 478.3, 486.4, 494.8, 503.2, and 511.8 K were used to obtain an exchange rate of 18%,[27] and the exchanges were attempted every 4 ps. The system was simulated for 1.6 $\mu$s per replica for a total time of 64 $\mu$s of simulation time.

Since there is not a reported experimental structure, single linkage clustering was used to find metastable states. The most populated cluster reproduces the expected features[22] (tight turn, face to edge contacts between the tryptophans and hydrogen bonds between the acetyl and the C-terminal), and thus, this structure is identified as the native state.

Histograms of $C_\alpha$-RMSD to the native state at different temperatures show two populations separated around 0.2 nm; Figure 5. The behavior with temperature suggests that using $C_\alpha$-RMSD = 0.2 nm is a good cutoff to separate folded from unfolded structures. At 275 K the peptide adopts a beta hairpin configuration with population in excess of 85%, consistent with experimental data.[22] With increasing temperature, the population of the folded state decreases and the melting temperature (temperature where the folded and unfolded states have equal probabilities) is found to be close to 330 K; Figure 5.

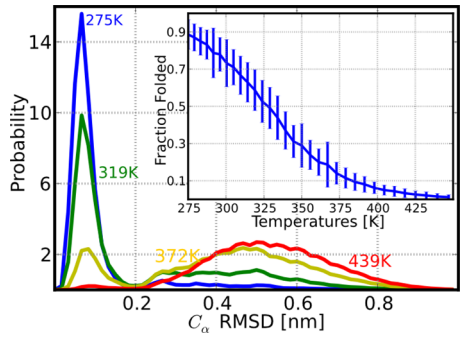



**Figure 5.** Histograms of $C_\alpha$-RMSD for various temperatures. Two different basins with a gap around 0.2 nm are found. The peak around 0.1 nm is destabilized with temperature while the wider one gets more populated suggesting a two-state thermodynamic behavior.

## ASSOCIATED CONTENT

### Supporting Information

Implementation and simulation details. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author

*To whom correspondence should be addressed E-mail: angel@rpi.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge support by the National Science Foundation MCB-1050966.

## REFERENCES

(1) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.
(2) Best, R. B. *Curr. Opin. Struct. Biol.* **2012**, *22*, 52−61.
(3) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120−127.
(4) Voelz, V. a.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 1526−1528.
(5) Yang, S.; Onuchic, J. N.; Garcia, A.; Levine, H. *J. Mol. Biol.* **2007**, *372*, 756−763.
(6) Johnson, M. E.; Hummer, G. *J. Phys. Chem. B* **2012**, *116*, 8573−8583.
(7) Henkelman, G.; Uberuaga, B.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9901−9904.
(8) Christoph Dellago, Peter G.; Bolhuis, F. S. C.; Chandler, D. *J. Chem. Phys.* **1997**, *108*, 1964−1977.
(9) Woolf, T. B. *Chem. Phys. Lett.* **1998**, *289*, 433−441.
(10) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187−199.
(11) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. *Computational molecular dynamics: challenges, methods, ideas*; Springer: New York, 1998; Vol. *4*, pp 39−65.
(12) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 5−8.
(13) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1997**, *350*, 205−23.
(14) Hansmann, U. H.; Okamoto, Y. *Curr. Opin. Struct. Biol.* **1999**, *9*, 177−83.
(15) Buchete, N.-V.; Hummer, G. *Phys. Rev. E* **2008**, *77*, 1−4.
(16) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. *J. Phys. Chem. B* **2009**, *113*, 11702−9.
(17) van der Spoel, D.; Seibert, M. *Phys. Rev. Lett.* **2006**, *96*, 1−4.
(18) Chodera, J. D.; Swope, W. C.; Noèl, F.; Prinz, J.-H.; Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2011**, *134*, 244107.

(19) Prinz, J.-H.; Chodera, J. D.; Pande, V. S.; Swope, W. C.; Smith, J. C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 244108.
(20) Pan, A. C.; Sezer, D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3432−40.
(21) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 24106.
(22) Kier, B. L.; Shu, I.; Eidenschink, L. A.; Andersen, N. H. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10466−10471.
(23) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(24) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Bioinformatics* **2006**, *725*, 712−725.
(25) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M.; et al. *J. Chem. Phys.* **1983**, *79*, 926.
(26) *The PyMOL Molecular Graphics System*; Schrödinger, LLC., New York, 2010.
(27) Garcia, A.; Herce, H.; Paschek, D. *Annu. Rep. Comput. Chem.* **2006**, *2*, 83−95.
(28) Garcia, A. *Phys. Rev. Lett.* **1992**, *68*, 2696−2700.
(29) Sindhikara, D.; Emerson, D.; Roitberg, A. *J. Chem. Theory Comput.* **2010**, *6*, 2804−2808.