

Investigations of α -helix β -sheet transition pathways in a miniprotein using the finite-temperature string method

Victor Ovchinnikov, and Martin Karplus

Citation: *The Journal of Chemical Physics* **140**, 175103 (2014);

View online: <https://doi.org/10.1063/1.4871685>

View Table of Contents: <http://aip.scitation.org/toc/jcp/140/17>

Published by the [American Institute of Physics](#)

Articles you may be interested in

[Free energy of conformational transition paths in biomolecules: The string method and its application to myosin VI](#)

The Journal of Chemical Physics **134**, 085103 (2011); 10.1063/1.3544209

[Revisiting the finite temperature string method for the calculation of reaction tubes and free energies](#)

The Journal of Chemical Physics **130**, 194103 (2009); 10.1063/1.3130083

[String method in collective variables: Minimum free energy paths and isocommittor surfaces](#)

The Journal of Chemical Physics **125**, 024106 (2006); 10.1063/1.2212942

[Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide](#)

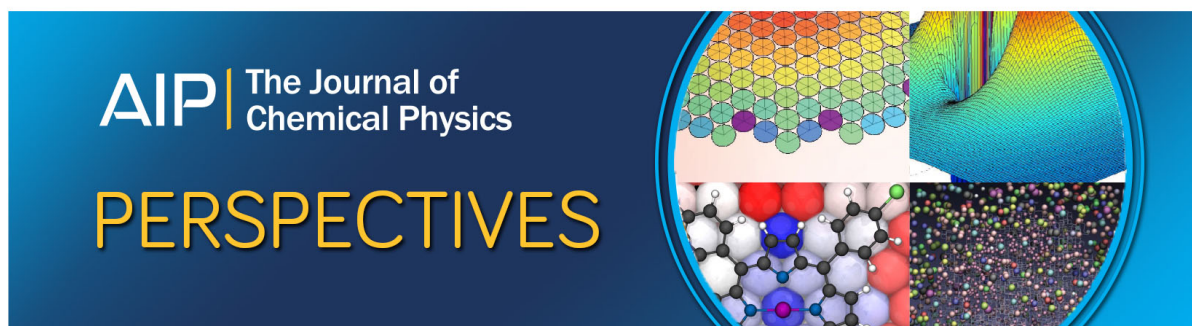
The Journal of Chemical Physics **123**, 134109 (2005); 10.1063/1.2013256

[Comparison of simple potential functions for simulating liquid water](#)

The Journal of Chemical Physics **79**, 926 (1998); 10.1063/1.445869

[Simulating rare events using a weighted ensemble-based string method](#)

The Journal of Chemical Physics **138**, 044105 (2013); 10.1063/1.4773892



Investigations of α -helix $\leftrightarrow\beta$ -sheet transition pathways in a miniprotein using the finite-temperature string method

Victor Ovchinnikov^{1,a)} and Martin Karplus^{1,2,b)}

¹Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA

²Laboratoire de Chimie Biophysique, ISIS, Université de Strasbourg, 67000 Strasbourg, France

(Received 23 December 2013; accepted 7 April 2014; published online 7 May 2014)

A parallel implementation of the finite-temperature string method is described, which takes into account the invariance of coordinates with respect to rigid-body motions. The method is applied to the complex α -helix $\leftrightarrow\beta$ -sheet transition in a β -hairpin miniprotein in implicit solvent, which exhibits much of the complexity of conformational changes in proteins. Two transition paths are considered, one derived from a linear interpolant between the endpoint structures and the other derived from a targeted dynamics simulation. Two methods for computing the conformational free energy (FE) along the string are compared, a restrained method, and a tessellation method introduced by E. Vanden-Eijnden and M. Venturoli [J. Chem. Phys. **130**, 194103 (2009)]. It is found that obtaining meaningful free energy profiles using the present atom-based coordinates requires restricting sampling to a vicinity of the converged path, where the hyperplanar approximation to the isocommittor surface is sufficiently accurate. This sampling restriction can be easily achieved using restraints or constraints. The endpoint FE differences computed from the FE profiles are validated by comparison with previous calculations using a path-independent confinement method. The FE profiles are decomposed into the enthalpic and entropic contributions, and it is shown that the entropy difference contribution can be as large as 10 kcal/mol for intermediate regions along the path, compared to 15–20 kcal/mol for the enthalpy contribution. This result demonstrates that enthalpic barriers for transitions are offset by entropic contributions arising from the existence of different paths across a barrier. The possibility of using systematically coarse-grained representations of amino acids, in the spirit of multiple interaction site residue models, is proposed as a means to avoid *ad hoc* sampling restrictions to narrow transition tubes. © 2014 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4871685>]

I. INTRODUCTION

Conformational transitions in proteins and DNA underlie many processes in biology, such as enzyme catalysis,¹ DNA replication,^{2,3} refolding of misfolded proteins,⁴ ATP synthesis and transport by molecular motors,^{5,6} muscle contraction,⁷ and multidrug resistance.⁸ The modeling of conformational changes in biomolecules remains one of the most challenging problems in computational biophysics because transitions in proteins often require micro- to milli-seconds, while typical atomistic MD simulations of proteins are restricted to hundreds of nanoseconds. Although special-purpose hardware for accelerating MD simulations to be able to study millisecond phenomena is becoming more readily available,^{9–11} many macromolecular processes remain beyond the reach of conventional MD.

To overcome the time-scale limitation of standard MD simulations, many enhanced sampling techniques have been developed, such as Umbrella Sampling (US),^{12,13} Metadynamics,¹⁴ Transition Path Sampling (TPS),^{15,16} Adaptive Biasing Force (ABF),^{17,18} Adiabatic MD,¹⁹ Targeted MD,²⁰ Temperature Accelerated (TA) MD,²¹ Milestoning,^{22–24} and the String Method.^{25,26} Most of these

methods require *a priori* specification of a reaction coordinate or of coarse-grained variables (e.g., US, metadynamics, ABF, TAM, string method in collective variables²⁶) which are not generally obvious for complicated molecular systems. TPS^{15,16} can refine an initial path connecting two metastable states without the need to specify a reaction coordinate or collective variables by using a random walk in the space of transition paths. However, because TPS relies on the generation of unbiased trajectories by MD, it can be very costly to apply to large biomolecules. Targeted MD requires only the specification of the endpoint structures, and generates transition paths within a prescribed simulation time using constraints^{20,27,28} or restraints.²⁹ However, because the dynamics in TMD are biased, the resulting transition paths can be far from equilibrium.³⁰ Further discussion of the various sampling methods can be found in recent reviews.^{31,32}

In this study, we focus on another approach, the finite-temperature string method^{25,33} (FTS), which was proposed to circumvent some of the requirements of the above sampling methods. FTS can be applied directly in Cartesian coordinates, without requiring additional coarse-grained variables or a reaction coordinate. FTS does not rely on the generation of unbiased trajectories, and is therefore expected to converge faster than TPS methods. In addition, free energy profiles of transitions can be calculated from the converged path in a straightforward way. FTS still requires the user to

^{a)}Electronic mail: ovchinnv@georgetown.edu

^{b)}Electronic mail: marci@tammy.harvard.edu

supply an initial path, which can be obtained from methods of varying computational complexity (e.g., interpolation, the nudged elastic band method,³⁴ conjugate peak refinement,³⁵ zero-temperature string (ZTS),³⁶ TMD). This choice can have a significant effect on the converged path in systems with rugged energy landscapes (e.g., see Sec. III in the present paper).

We describe the application of the FTS to the α -helix $\leftrightarrow\beta$ -sheet transition in a G-protein peptide.³⁷ This model system possesses much of the complexity of realistic proteins, and is therefore used here to assess the suitability of using the FTS to study arbitrary conformational changes. In Sec. II, we summarize the theory underlying the FTS, present a parallel implementation in rigid-body-invariant coordinates, and describe two techniques for computing the free energy profile from the string. In Sec. III, we describe the transition paths resulting from the application of the FTS to the G-protein peptide, along with the profiles of free energy, enthalpy, and entropy. In Sec. IV, we discuss the limitations of applying the FTS in the given coordinates, and in Sec. V, we provide a summary and directions for future work.

II. METHODS

A. Summary of the finite-temperature string method

A rigorous derivation of the FTS method from transition path theory (TPT)³⁸ has been given in Refs. 25 and 38–40. Briefly, in TPT the central object is the committor function q , which corresponds to the solution to the backward Kolmogorov equation in probability theory.⁴¹ Given a continuous time-evolving stochastic system whose state is described by $\mathbf{x}(t)$ and a definition of a *reactants* set (X_A) and a *products* set (X_B), the committor $q(\mathbf{y})$ corresponds to the probability that a system in state \mathbf{y} will enter (*commit* to) the product set before reaching the reactant set. It is clear that the committor function is a natural measure of the reaction progress, and it has been used as a reaction coordinate in TPS simulations of various phenomena.^{15,16,42–45} In the remainder of the paper, we will not make a distinction between the terms “committor” and “reaction coordinate.”

The FTS can be derived from the backward Kolmogorov equation for Langevin dynamics on a potential energy surface ($U(\mathbf{x})$) in the high-friction limit³⁹ under the assumption that the isosurface $q = q_0$ of the committor can be represented by a hyperplane, i.e.,

$$|\nabla q(\mathbf{x})| \delta(q(\mathbf{x}) - q_0) d\mathbf{x} = \delta(\mathbf{n}(q_0) \cdot [\mathbf{x} - \boldsymbol{\phi}(q_0)]) d\mathbf{x}, \quad (1)$$

where δ is the Dirac distribution, $|\nabla q(\mathbf{x})|$ is the Jacobian factor, $\mathbf{n}(q_0)$ is a unit normal to the hypersurface at $q = q_0$, and $\boldsymbol{\phi}$ is constrained to satisfy the equation

$$\begin{aligned} \boldsymbol{\phi}(q_0) &= Z(q_0)^{-1} \int \mathbf{x} e^{-\beta U(\mathbf{x})} \delta(\mathbf{n}(q_0) \cdot [\mathbf{x} - \boldsymbol{\phi}(q_0)]) d\mathbf{x} \\ &= \langle \mathbf{x} \rangle_{P_{q_0}}. \end{aligned} \quad (2)$$

$Z(q_0)$ is the partition function corresponding to the isosurface and the subscript P_{q_0} denotes canonical sampling restricted to the hyperplane at $q = q_0$. Thus, $\boldsymbol{\phi}(q_0)$ is the *average* position in the hyperplane associated with q_0 , and the curve $\boldsymbol{\phi}$, defined

as the *string*, is the *average* reaction path. Provided that the variance of the quantity $(\mathbf{n}(q_0) \cdot [\mathbf{x} - \boldsymbol{\phi}(q_0)])$ is not too large, $\boldsymbol{\phi}$ also represents the dominant reaction path. With the definitions in Eqs. (1) and (2) it can be shown that, for Langevin dynamics in the high-friction limit, the hyperplane normal is parallel to the tangent vector to the string,³⁹

$$\mathbf{n}(q_0) \parallel \boldsymbol{\phi}'(q_0), \quad (3)$$

i.e., within the accuracy of the hyperplane approximation, the isosurfaces of the reaction coordinate can be obtained from the string directly.

The FTS is an iterative algorithm for solving a discretized version of Eqs. (2) and (3).^{25,33,40} From an approximation to the string at iteration i ($\boldsymbol{\phi}^i$) one obtains n^i using Eq. (3), which permits computing $\boldsymbol{\phi}^{i+1}$ from Eq. (2). The procedure is repeated until $\boldsymbol{\phi}^i$ does not change. The free energy can then be obtained by thermodynamic integration of the free energy derivatives sampled on the hyperplanes,^{25,33} or by sampling a Voronoi tessellation.⁴⁶ The FTS algorithm with Voronoi tessellations has also been applied to compute probability distributions and reactive fluxes in nonequilibrium steady-state problems.^{47–49}

The original publications of the FTS^{25,40} did not explicitly consider the rigid-body invariance of systems that are free to translate and rotate in space, which includes most macromolecules simulated by molecular dynamics (MD). In the first application of the string method to the model biomolecule alanine dipeptide,³³ simple harmonic restraints were applied to keep the molecule from rotating or translating in space. Other applications of the string method^{50,51} used “best-fit” transformations of the absolute coordinates to a frame of reference associated with a particular structure^{52–55} (e.g., an xray or an energy-minimized structure of a protein) to represent the string of replicas. This approach dates back to the chain-of-states methods^{56,57} such as the nudged elastic band,^{34,58} or path functional minimization.⁵⁹ It has been shown that the best-fitting is equivalent to the application of the Eckart conditions.^{30,60–63} Vanden-Eijnden and Venturoli⁴⁶ described a tessellation approach to remove degeneracy due to translation and rotation using a minimal distance metric, computed after the instantaneous simulation coordinates are best-fit onto a string replica. While the tessellation approach appears technically simple, it is based on constrained dynamics, and requires modifying the MD integrator. It is therefore not portable between different MD programs, or even between different integrators within the same MD program. To address this difficulty, Maragliano, Vanden-Eijnden, and Roux⁶⁴ presented a method to sample a Voronoi tessellation using one-sided restraint dynamics, and applied it to solvated alanine dipeptide in collective variable space.

The present implementation of the FTS described below combines the idea of the degenerate distance metric of Vanden-Eijnden and Venturoli⁴⁶ with the idea of using restraints to compute one-dimensional free energy profiles.⁶⁴ The approach uses restrained dynamics with a harmonic potential centered on each hypersurface, and is therefore expected to be easily portable between MD programs. Both the restrained sampling method and the tessellation method are used to calculate the free energy profiles. The results

are compared in Sec. III C. An extension of the tessellation sampling approach that does not involve modifications to the integrator is the weighted ensemble (WE) string method.⁶⁵ In this method, sampling constraints are not enforced explicitly during the MD simulation. Instead, simulations corresponding to different cells in the tessellation are reinitialized periodically after a prescribed time interval τ to ensure uniform sampling of the tessellation. The efficiency of the WE method appears to be sensitive to the choice of the parameter τ , which should be greater than the trajectory decorrelation time of the system.⁶⁵ However, a comparison between the WE method and the restrained method presented here is beyond the scope of this study.

There are other methods that account for the degeneracy due to rigid body motion by re-expressing atomic coordinates in the reference frame of the moment-of-inertia tensor.^{66,67} These methods require additional care in systems for which the eigenvalues of the tensor are not well separated to avoid discontinuities in the eigenvectors during MD simulation.

B. Rotation-invariant FTS

In this subsection, we describe a **rigid-body-invariant implementation of the FTS** based on Eqs. (2) and (3). Let $\hat{r}^* = \{r_1^*, \dots, r_N^*\}$ be set of N coordinate triplets of the atoms that undergo a conformational transition, and $w_i = m_i / \sum_{j=1}^N m_j$, where m_i are the atomic masses. Center-of-mass (COM) coordinates are defined by $r_j = r_j^* - \sum_{k=1}^N w_k r_k^*$. Henceforth, we assume that all coordinates, including points along the string, have been shifted to the corresponding COMs. Defining a parameter $\alpha \in [0, 1]$, we represent the string as $\hat{\phi}(\alpha) = \hat{\phi}(q_0(\alpha))$. The parametrization $\hat{\phi}(\alpha)$ rather than $\hat{\phi}(q_0)$ is used because the values of the committor q_0 can be obtained from the string only after it is converged.²⁵

By analogy with Eqs. (1) and (3), we assume that the isosurface (or level set) of the committor function at $q = q_0(\alpha)$ can be approximated by hypersurfaces defined by

$$\hat{\phi}'(\alpha) \cdot [A^T \hat{r} - \hat{\phi}(\alpha)] = 0, \quad (4)$$

where $A(\hat{r}, \hat{\phi})$ with $A^T A = I$ is a rotation matrix for which the squared norm

$$\begin{aligned} \|A^T \hat{r} - \hat{\phi}(\alpha)\|^2 &= \|\hat{r} - A\hat{\phi}(\alpha)\|^2 \\ &= [\hat{r} - A\hat{\phi}(\alpha)] \cdot [\hat{r} - A\hat{\phi}(\alpha)] \\ &\equiv \sum_{j=1}^N w_j |\hat{r}_j - A\hat{\phi}_j(\alpha)|^2 \end{aligned} \quad (5)$$

is minimized over all possible rotations. A corresponds to the “best-fit” superposition of the image $\phi(\alpha)$ onto the coordinates \hat{r} .⁶⁸ (Conversely, A^T corresponds to the best-fit of \hat{r} onto $\phi(\alpha)$, i.e., $A(\hat{\phi}, \hat{r}) = A^T(\hat{r}, \hat{\phi})$.) For brevity, we will sometimes omit the explicit dependence of A on \hat{r} and $\hat{\phi}(\alpha)$. Invariance with respect to rotations of \hat{r} is implicit in Eq. (4) because for any \hat{r} that satisfies Eq. (4) with some $A = A^*$, and any rigid rotation operation R , $R\hat{r}$ will also satisfy Eq. (4) with $A = RA^*$. The relationship between Eq. (4) and the distance metric used by Vanden-Eijnden and Venturoli⁴⁶ is made clear in Sec. II D

in which the tessellation approach for computing free energy profiles is reviewed.

The connection between Eq. (4) and the original hyperplanar ansatz in Eq. (1) can be seen by considering the change of variables $\hat{r}^* \mapsto A^T \hat{r}$ and the associated mass-weighted metric tensor

$$\begin{aligned} M_{i,j}^{a,b}(\hat{\phi}(\alpha)) \\ = \sum_{c=1}^3 \sum_{k=1}^N \frac{1}{m_k} \left\langle \frac{\partial [A^T(\hat{r}, \hat{\phi}) \hat{r}_i^a]}{\partial r_k^{*c}} \frac{\partial [A^T(\hat{r}, \hat{\phi}) \hat{r}_j^b]}{\partial r_k^{*c}} \right\rangle_{A^T \hat{r} = \hat{\phi}(\alpha)}. \end{aligned} \quad (6)$$

It was shown in Ref. 26 that, for an arbitrary change of variables $\hat{r}^* \mapsto \theta(\hat{r}^*)$, the vector normal to the isocommittor surface transforms according to $\hat{n}_{\hat{r}^*}(\alpha) \mapsto M(\theta(\alpha)) \hat{n}_{\theta}$. The use of Eq. (4) is therefore justified when $M_{i,j}^{a,b} \approx \delta_{ab} \delta_{ij} m_j^{-1}$, i.e., when M is approximately equal to the diagonal inverse mass matrix. The metric tensor was computed for two systems, alanine dipeptide in vacuum, and a 16-residue β -hairpin peptide in implicit solvent. The details and results of the calculations are provided in the Appendix. The norm of the difference tensor $D_{i,j}^{a,b} = M_{i,j}^{a,b} - \delta_{ab} \delta_{ij} m_j^{-1}$ defined as

$$D = \frac{1}{(3N)^2} \sum_{a,b=1}^3 \sum_{i,j=1}^N |D_{i,j}^{a,b}| (m_i m_j)^{1/2}, \quad (7)$$

was $\simeq 0.05$ for the alanine dipeptide (which has 22 atoms in the all-atom representation, but $N = 10$ because hydrogen atoms were not treated as independent particles in the FTS), and $\simeq 0.005$ for a 16-residue β -hairpin peptide (which has 247 atoms, but only $N = 107$ atoms were selected as independent particles; see Sec. III). Equation (4) is therefore a good approximation to the original condition Eq. (1), and is used in the present method to find an iterative solution to the string $\hat{\phi}$, up to a rigid-body rotation, i.e.,

$$\hat{\phi}(\alpha) = \langle A^T \hat{r} \rangle_{\perp \hat{\phi}(\alpha)}. \quad (8)$$

We note that the use of the best fitting operation in Eq. (4) is equivalent to the construction of the Eckart reference frame^{60,61,69–71} for the approximate separation of rotational and vibrational motions in semi-rigid molecules. For the present calculations using the FTS it was found that some instantaneous molecular configurations ($A^T \hat{r}$) are distributed broadly around the string ($\hat{\phi}$), with the distances in Eq. (5) ~ 3 Å (which underscores the importance of configurational entropy to the free energy of proteins in solution; see Sec. III). Although such large values may invalidate the assumption of small vibrational displacements used in the Eckart construction, the relatively small norm of the difference tensor justifies the use of the Eckart frame.

Because practical use of the FTS requires a discrete representation of the string by a finite set of points (called “images” below), an additional condition must be imposed to maintain a prescribed numerical resolution. In the present case, we use

$$\|(\hat{A}\hat{\phi})'_{|\hat{r}=\hat{\phi}}\|' = 0, \quad (9)$$

which corresponds to parametrization by arclength.⁴⁰ The difference between Eq. (9) and the original constraint equation

$(\|\hat{\phi}'\|' = 0)$ ⁴⁰ becomes nonzero for the discretized version of the string method, as can be seen by rewriting Eq. (9) as a finite-difference limit (and using $A(\hat{\phi}, \hat{\phi}) = I$),

$$\begin{aligned} & \lim_{d\alpha \rightarrow 0} \|A(\hat{\phi}(\alpha), \hat{\phi}(\alpha + d\alpha))\hat{\phi}(\alpha + d\alpha) \\ & \quad - A(\hat{\phi}(\alpha), \hat{\phi}(\alpha))\hat{\phi}(\alpha)\|/d\alpha \\ &= \lim_{d\alpha \rightarrow 0} \|A(\hat{\phi}(\alpha), \hat{\phi}(\alpha + d\alpha))\hat{\phi}(\alpha + d\alpha) - \hat{\phi}(\alpha)\|/d\alpha \\ &= L. \end{aligned} \quad (10)$$

Thus, Eq. (9) fixes the distances between adjacent string images at $Ld\alpha$, where L is string length, after each string image is best-fit to the preceding image.

Defining the unit normal $\hat{n}(\alpha) = \hat{\phi}'(\alpha)/\|\hat{\phi}'(\alpha)\|$, the canonical average of a quantity $a(\hat{r})$ restricted to the hypersurface $q(\hat{r}) = q_0(\alpha)$ (henceforth denoted by S_α) is

$$\begin{aligned} \langle a \rangle_{S_\alpha} &\equiv Z(\alpha)^{-1} \\ &\times \int_{\mathbb{R}^{3N}} a(\hat{r}) e^{-\beta U(\hat{r})} \delta([\hat{n}(\alpha)] \cdot [A^T \hat{r} - \hat{\phi}]) |\nabla(A^T \hat{r})|, \end{aligned} \quad (11)$$

where $U(\hat{r})$ is the potential energy, $Z(\alpha)$ is the partition function

$$Z(\alpha) = \int_{\mathbb{R}^{3N}} e^{-\beta U(\hat{r})} \delta([\hat{n}(\alpha)] \cdot [A^T \hat{r} - \hat{\phi}]) |\nabla(A^T \hat{r})|, \quad (12)$$

and $|\nabla(A^T \hat{r})|$ is the Jacobian associated with the change of variables $\hat{r}^* \mapsto A^T \hat{r}$. The Jacobian can be computed from the (unweighted) metric tensor determinant as $(\det M)^{1/2}$, after the six rigid body degrees of freedom are accounted for, as done in the Appendix. Using the β -hairpin system as an example, it is shown that the Jacobian factor is nearly constant along the string, varying within about 3% of the average value. This observation is consistent with the fact that M is close to the identity, and suggests that the Jacobian can be replaced by a constant, which disappears from the configurational averages in Eq. (11). The resulting partition function $Z(\alpha)$ excludes translational and rotational contributions, which can be included approximately *a posteriori* (see Sec. IV). The small deviations of the Jacobian and the metric tensor from unity are caused by the inexact separability of rotational and internal degrees of freedom.^{60,71-73}

To emphasize the fact that the molecular system confined to the hypersurface described by Eq. (4) can rotate freely, we can switch to the (instantaneous) frame of \hat{r} by applying $I = A^T A$ inside Eq. (4) to obtain

$$\begin{aligned} 0 &= \hat{\phi}'(\alpha) \cdot [A^T \hat{r} - \hat{\phi}(\alpha)] = A\hat{\phi}'(\alpha) \cdot [\hat{r} - A\hat{\phi}(\alpha)] \\ &= (A\hat{\phi})'(\alpha) \cdot [\hat{r} - A\hat{\phi}(\alpha)], \end{aligned} \quad (13)$$

where the third equality follows from the fact that $\hat{\phi}^T(\alpha)A^T A\hat{\phi}(\alpha)$ and $\hat{\phi}^T(\alpha)A^T \hat{r}$ are both zero. (The first term is zero because $A^T A$ is skew-symmetric; that the second term is zero can be seen by setting to zero the variation of the best-fit condition [Eq. (5)] with respect to the infinitesimal rotation $\epsilon A^T A$, as done in Ref. 30.) Equation (13) shows that the hypersurface constraints (or restraints) are applied effectively after the string image $\hat{\phi}$ is optimally rotated to overlap with

the instantaneous coordinates \hat{r} . To compute the free energy associated with the hyperplane at α , i.e.,

$$F(\alpha) = -\beta^{-1} \log Z(\alpha), \quad (14)$$

we define the auxiliary potential

$$\begin{aligned} U_k(\hat{r}) &= U(\hat{r}) + \frac{k}{2} ([\hat{n}(\alpha)] \cdot [A^T \hat{r} - \hat{\phi}])^2 \\ &= U(\hat{r}) + \frac{k}{2} ([A\hat{n}(\alpha)] \cdot [\hat{r} - A\hat{\phi}])^2. \end{aligned} \quad (15)$$

The corresponding partition function and free energy are

$$Z_k(\alpha) = \int_{\mathbb{R}^{3N}} e^{-\beta U_k(\hat{r})}, \quad (16)$$

$$F_k(\alpha) = -\beta^{-1} \log Z_k(\alpha). \quad (17)$$

Up to errors of order $O(1/k)$,²⁶

$$\langle a \rangle_{S_\alpha} \approx Z_k^{-1} \int_{\mathbb{R}^{3N}} a(\hat{r}) e^{-\beta U_k(\hat{r})}. \quad (18)$$

The right-hand side of Eq. (18) can be estimated from a molecular dynamics simulation using the auxiliary potential $U_k(\hat{r})$, i.e.,

$$\langle a \rangle_{S_\alpha} \approx \frac{1}{T} \int_{t_0}^{t_0+T} a(\hat{r}(t)) dt. \quad (19)$$

The free energy $F(\alpha)$ can be computed using thermodynamic integration (TI) and Eq. (19) with $a = \partial F_k / \partial \alpha$, i.e.,

$$F(\alpha) - F_0 = \int_0^\alpha \frac{dF}{d\alpha} d\alpha \approx \int_0^\alpha \frac{dF_k}{d\alpha} d\alpha. \quad (20)$$

To evaluate $dF_k/d\alpha$, we first differentiate the scalar product $[A\hat{n}(\alpha)] \cdot [\hat{r} - A\hat{\phi}(\alpha)]$,

$$\begin{aligned} & ([A\hat{n}(\alpha)] \cdot [\hat{r} - A\hat{\phi}(\alpha)])' \\ &= [A\hat{n}(\alpha)]' \cdot [\hat{r} - A\hat{\phi}(\alpha)] - [A\hat{n}(\alpha)] \cdot [A\hat{\phi}(\alpha)]' \\ &= A\hat{n}'(\alpha) \cdot [\hat{r} - A\hat{\phi}(\alpha)] + A'\hat{n}(\alpha) \cdot \hat{r} - L \\ &= [A\hat{n}(\alpha)]' \cdot \hat{r} - \hat{n}'(\alpha) \cdot \phi - L, \end{aligned} \quad (21)$$

where we have defined a curvature projection $C(\alpha)$ and used parametrization by arclength Eq. (9). Previous implementations of the FTS^{25,33} used *localization* assumptions, e.g., $|C(\alpha)| \ll L$, to neglect the curvature term in the force computation. The assumption is justified if either the curvature of the string, or the width of the transition tube are small. We have found that the curvature term gives rise to contributions of 1 kcal/mol to the free energy profile for the alanine dipeptide in vacuum (see the supplementary material⁷⁴). Therefore, the term is retained in the present calculations.

Using Eq. (21), and defining $\xi \equiv C(\alpha)/L$, we have

$$\begin{aligned} \frac{dF_k}{d\alpha} &= -\frac{kL}{Z_k(\alpha)} \\ &\times \int_{\mathbb{R}^{3N}} (1 - \xi) [A\hat{n}(\alpha)] \cdot [\hat{r} - A\hat{\phi}(\alpha)] e^{-\beta U_k(\hat{r})} \\ &\approx -\frac{kL}{T} \int_{t_0}^{t_0+T} (1 - \xi) [A\hat{n}(\alpha)] \cdot [\hat{r}(t) - A\hat{\phi}(\alpha)] dt, \end{aligned} \quad (22)$$

where the last integral represents an average over a MD trajectory of duration \mathcal{T} performed with the potential $U_k(\hat{r})$.

C. Discretization of the FTS

In this section, we give discretized analogs of Eqs. (20)–(22) needed for practical FTS calculations. Let M correspond to the number of points in the discretization of the string. The discrete string then consists of M “images” indexed by M values of the parameter α , i.e., $\hat{\phi} = \{\hat{\phi}_1 \equiv \hat{\phi}(\alpha_1), \dots, \hat{\phi}_M \equiv \hat{\phi}(\alpha_M)\}$. From Eq. (13), the hypersurface $S_i \equiv S_{\alpha_i}$ corresponding to the string image $\hat{\phi}_i$ is defined by

$$(A\hat{\phi})'_i \cdot [\hat{r} - A_i\hat{\phi}_i] = 0. \quad (23)$$

The reason for using the form in Eq. (23) (i.e., differentiating A along with $\hat{\phi}$) is that the free energy can then be computed without aligning any of the string images to each other, as will be seen below. However, a discretization based on Eq. (4) gives essentially the same results, provided that care is taken to superpose each string image onto the preceding image prior to computing the free energy. For an internal string image $\hat{\phi}_i$ (i.e., with $1 < i < M$), we define a finite difference approximation to the unit normal to S_i using the neighboring images $\hat{\phi}_{i\pm 1}$,

$$\mathcal{N}(\alpha_i) \approx \frac{A_{i+1}\hat{\phi}_{i+1} - A_{i-1}\hat{\phi}_{i-1}}{\|A_{i+1}\hat{\phi}_{i+1} - A_{i-1}\hat{\phi}_{i-1}\|}, \quad (24)$$

where $A_i = A_i(\hat{r}, \hat{\phi}_i)$ is the best-fit rotation matrix for the image $\hat{\phi}_i$. The unit normals associated with the endpoint string images $\hat{\phi}_1$ and $\hat{\phi}_M$ are defined using one-sided analogs of Eq. (24), i.e.,

$$\begin{aligned} \mathcal{N}(\alpha_1) &\approx \frac{A_2\hat{\phi}_2 - A_1\hat{\phi}_1}{\|A_2\hat{\phi}_2 - A_1\hat{\phi}_1\|}, \\ \mathcal{N}(\alpha_M) &\approx \frac{A_M\hat{\phi}_M - A_{M-1}\hat{\phi}_{M-1}}{\|A_M\hat{\phi}_M - A_{M-1}\hat{\phi}_{M-1}\|}. \end{aligned} \quad (25)$$

To second-order accuracy in the distance between string images $L/(M-1)$, where L is the string length, we make the approximation

$$A_i\hat{\phi}_i \approx \frac{1}{2}(A_{i-1}\hat{\phi}_{i-1} + A_{i+1}\hat{\phi}_{i+1}). \quad (26)$$

Using Eqs. (24) and (26), Eq. (23) can be written as

$$\begin{aligned} \epsilon_i &\equiv \mathcal{N}_i \cdot [\hat{r} - A_{i-1}\hat{\phi}_{i-1}] \\ &= \|A_{i+1}\hat{\phi}_{i+1} - A_{i-1}\hat{\phi}_{i-1}\|/2 \approx L/(M-1), \quad 1 < i < M, \end{aligned} \quad (27)$$

where we have defined the variables ϵ_i to simplify notation below. The hypersurfaces associated with the endpoint string images $\hat{\phi}_1$ and $\hat{\phi}_M$ are defined using one-sided analogs of Eq. (27),

$$\epsilon_1 \equiv \mathcal{N}_1 \cdot [\hat{r} - A_1\hat{\phi}_1] = 0, \quad (28)$$

$$\epsilon_M \equiv \mathcal{N}_M \cdot [\hat{r} - A_{M-1}\hat{\phi}_{M-1}] = L/(M-1). \quad (29)$$

Introducing the constants

$$\epsilon_i^0 \equiv \begin{cases} 0, & i = 1 \\ L/(M-1), & 1 < i \leq M \end{cases} \quad (30)$$

the finite-difference approximation to the canonical average in Eq. (11) at $\alpha = \alpha_i$ is

$$\begin{aligned} \langle a \rangle_{S_i} &= (1/Z^i) \int_{\mathbb{R}^{3N}} a(\hat{r}) e^{-\beta U(\hat{r})} \delta(\epsilon_i(\hat{r}) - \epsilon_i^0) \\ &\approx (1/Z_k^i) \int_{\mathbb{R}^{3N}} a(\hat{r}) e^{-\beta U_k^i(\hat{r})}, \end{aligned} \quad (31)$$

where $U_k^i(\hat{r}) = U(\hat{r}) + (k/2)(\epsilon_i(\hat{r}) - \epsilon_i^0)^2$, and Z^i and Z_k^i are normalization constants.

The curvature correction $\xi(\alpha)$ is computed for internal string images with the help of the approximation

$$\begin{aligned} (\mathcal{N})'_i &\approx (M-1) \\ &\times \left\{ \frac{A_{i+1}\hat{\phi}_{i+1} - A_i\hat{\phi}_i}{\|A_{i+1}\hat{\phi}_{i+1} - A_i\hat{\phi}_i\|} - \frac{A_i\hat{\phi}_i - A_{i-1}\hat{\phi}_{i-1}}{\|A_i\hat{\phi}_i - A_{i-1}\hat{\phi}_{i-1}\|} \right\} \\ &\approx \frac{(M-1)^2}{L} \{A_{i+1}\hat{\phi}_{i+1} - 2A_i\hat{\phi}_i + A_{i-1}\hat{\phi}_{i-1}\}, \end{aligned} \quad (32)$$

where we used $\|A_{i+1}\hat{\phi}_{i+1} - A_i\hat{\phi}_i\| \approx Ld\alpha$ for all $0 < i < M$ with $d\alpha \approx 1/(M-1)$. For the endpoint string images, the curvature correction was set to zero. From Eqs. (20) and (22) and the trapezoidal rule, the free energy associated with the hypersurface S_j is computed as

$$F(\alpha_j) \approx F_0 + \frac{1}{M-1} \sum_{i=1}^j \frac{1}{2} \left(\frac{dF_k^i}{d\alpha} + \frac{dF_k^{i-1}}{d\alpha} \right), \quad (33)$$

where

$$\frac{\partial F_k^i}{\partial \alpha} = -\frac{kL}{\mathcal{T}} \int_{t_0}^{t_0+\mathcal{T}} (1 - \xi_i)(\epsilon_i(\hat{r}) - \epsilon_i^0) dt. \quad (34)$$

D. Free energy computation by tessellations

As mentioned in the Introduction, the free energy can also be evaluated by constructing a tessellation of the configurational space based on the discretized string.⁴⁶ The tessellation can be sampled using either constrained^{46,67,75} or restrained⁶⁴ dynamics, and the free energy of the tessellation can be obtained from the steady-state solution of the corresponding Master equation.^{46,64,67,75} For the present calculations, the original constrained method was implemented, as follows. To each string image $\hat{\phi}_i$ we assign a cell B_i , defined by

$$B_i = \{\hat{r} : \|\hat{r} - A\hat{\phi}_i\| < \|\hat{r} - A\hat{\phi}_j\|, \forall j \neq i\}. \quad (35)$$

The equation for the boundary between the adjacent cells B_i , B_{i+1} , can be obtained by setting $j = i+1$, and rewriting the equality

$$\begin{aligned} \|\hat{r} - A\hat{\phi}_{i+1}\|^2 &= \|\hat{r} - A\hat{\phi}_i\|^2, \\ \|A\hat{\phi}_{i+1}\|^2 - \|A\hat{\phi}_i\|^2 &= 2\hat{r} \cdot [A\hat{\phi}_{i+1} - A\hat{\phi}_i], \\ [\hat{r} - [A\hat{\phi}_{i+1} + A\hat{\phi}_i]/2] \cdot [A\hat{\phi}_{i+1} - A\hat{\phi}_i] &= 0, \end{aligned} \quad (36)$$

where the last line corresponds to the second-order central finite-difference approximation to Eq. (23) at $\alpha = (\alpha_{i+1} + \alpha_i)/2$. Equation (36), therefore, shows that boundaries between adjacent cells approximate isocommittor surfaces, as noted before.^{46,67} The free energy of the tessellation can be computed by performing M independent MD simulations in the canonical ensemble, with the simulated systems constrained to remain inside the corresponding B_i . In the present calculations, the simple Brünger-Brooks-Karplus (BBK) Langevin Dynamics integrator⁷⁶ is used (which is the default integrator in CHARMM⁷⁷ and in NAMD⁷⁸), with the following reflection rule to enforce the constraint:

$$\begin{aligned} \text{if } \hat{r}_i(t) \notin B_i, \text{ then} \\ \Delta \hat{r}_i(t) = -\Delta \hat{r}_i(t - \delta t) // \text{reflect}, \\ \hat{r}_i(t + \delta t) = \hat{r}_i(t) + \Delta \hat{r}_i(t) // \text{propagate}, \end{aligned} \quad (37)$$

where the term $\Delta \hat{r}_i(t)$ represents the propagation of the particle i over the period $[t, t + \delta t]$. The free energy of the tessellation can be obtained from the time record of reflections, as described before.⁷⁵

E. Limiting the width of the transition tube

A central assumption of the FTS (Eq. (1)) is that isocommittor surfaces can be approximated by hyperplanes. While the orientation of the hyperplane (i.e., the normal \hat{n}) is derived rigorously in transition path theory,^{38,39} the distance from the string over which the hyperplane representation is reasonably accurate is unknown, and dependent on the details of the system. Even if a higher-order expansion of the isocommittor were available (e.g., with local curvature information), practical calculations would quickly become prohibitive in high-dimensional spaces because they would require multiplications of tensors of progressively higher rank. For these reasons it may be useful to impose additional restraints (or constraints) on the simulation system to keep it sufficiently close to the string, so that the hyperplane approximation is accurate. (This approach neglects some entropic contributions to the free energy profiles and therefore can lead to an increase in the calculated free energy barriers.)

We define the following optional restraint potential $V_\kappa(\hat{r})$ that limits the width of the transition tube to a prescribed value η^0 ,

$$V_\kappa^i(\hat{r}) = \frac{\kappa}{2} \min\{(\eta_i - \eta^0), 0\}^2, \quad (38)$$

where $\eta_i = \|v_i^\perp\|$ and

$$v_i^\perp = [\hat{r} - A_{i-1}\hat{\phi}_{i-1}] \cdot [I - (A\hat{n})_i(A\hat{n})_i^T] \quad (39)$$

is the component of $[\hat{r} - A_{i-1}\hat{\phi}_{i-1}]$ orthogonal to the string at $\hat{\phi}_i$. Using Eqs. (24) and (26), and the corresponding one-sided definitions for the endpoint images (Eqs. (28) and (29)), η_i and its derivatives can be computed with very little effort once ϵ_i and its derivatives are known (see the Appendix).

Tube width restrictions can also be implemented in the sampling of the tessellation defined in Subsection II D by

modifying the reflection rule (Eq. (37)),

$$\begin{aligned} \text{if } \hat{r}_i(t) \notin B_i \text{ or } \eta_i(t) > \eta^0, \text{ then} \\ \Delta \hat{r}_i(t) = -\Delta \hat{r}_i(t - \delta t) // \text{reflect}, \\ \hat{r}_i(t + \delta t) = \hat{r}_i(t) + \Delta \hat{r}_i(t) // \text{propagate}. \end{aligned} \quad (40)$$

Tests of the effect of these restrictions can be made by varying the prescribed width parameter η_0 in Eq. (38), as done in Sec. III.

F. Parallel implementation

The FTS requires an initial string $\hat{\phi}_i^0$, $i \in \{1, \dots, M\}$, which can be obtained from, e.g., a minimum energy path (MEP) or a biased dynamics trajectory (see Sec. III). To each image $\hat{\phi}_i$ we assign a separate CPU group \mathcal{I} and a complete all-atom MD simulation system, denoted by \hat{r}_i , to be used for calculating the expectation values in Eq. (31).

Each CPU group is provided with the (neighbor) image(s) $\hat{\phi}_{i\pm 1}^0$ in addition to $\hat{\phi}_i^0$. A running average of \hat{r}_i is initialized as $\hat{\psi}_i^0 = \hat{\phi}_i^0$. Starting at iteration $n = 0$, the following sequence of steps is repeated until the string is converged:

1. *Sampling.* Using the potential U_k^i (and V_κ^i , if limiting the transition tube width is desired), each CPU group \mathcal{I} performs a MD simulation with a time step Δt (typically 1–2 fs) for a prescribed simulation duration $\mathcal{T}^* = \mathcal{T}_{eq} + \mathcal{T}$ (where \mathcal{T} is defined in Eq. (19), and \mathcal{T}_{eq} is an equilibration period during which samples are discarded). The running average $\hat{\psi}_i^j$ is updated after each m successive steps:

$$\hat{\psi}_i^j = \frac{j-1}{j} \hat{\psi}_i^{j-1} + \frac{1}{j} A^T(\hat{r}, \hat{\psi}_i^{j-1}) \hat{r}. \quad (41)$$

2. *Reparametrization.* The running average $\hat{\psi}_i^j$ is reparametrized to enforce Eq. (9) using linear interpolation, as described in Refs. 26, 36, and 67, except that each string image $\hat{\phi}_{i>1}$ is first superposed onto the neighbor image $\hat{\phi}_{i>0}$ consistent with the parametrization constraint (Eq. (9)). Communication between different CPU groups is required for this step. As in the previous implementations,^{26,36,67} the reparametrization procedure is approximate and can be performed iteratively to achieve arbitrary accuracy. The reparametrization error is measured by the quantity

$$err = \max_{1 \leq i \leq M} err_i \equiv \left| \|\hat{\phi}_{i-1} - \hat{\phi}_i\| / \|\hat{\phi}_i - \hat{\phi}_{i+1}\| - 1 \right|, \quad (42)$$

which is computed after superposition of each string image onto the preceding image, as prescribed by Eq. (10). In the present calculations, once the string reached steady state (see *Convergence Test* below), err was around 0.01 after a single iteration. With the distance between the adjacent images of about 0.25 Å, the 1% error corresponds to uncertainties in the string image positions of only ~ 0.0025 Å.

3. *Update.* The string images are *updated* from the running average

$$\hat{\phi}_i^{n+1} = \hat{\psi}_i^j. \quad (43)$$

4. *Convergence test.* Convergence of the string is checked by monitoring the norm

$$\rho^{n+1} = \left[\sum_{i=1}^M \|\hat{\phi}_i^0 - A(\hat{\phi}_i^0, \hat{\phi}_i^{n+1})\hat{\phi}_i^{n+1}\|^2 \right]^{1/2}, \quad (44)$$

which corresponds to the root-mean-square distance through which the string images have moved at iteration $n + 1$. The loop is terminated after ρ^n reaches a plateau. Before returning to step 1, the neighboring images $\hat{\phi}_{i\pm 1}^{n+1}$ are communicated to each CPU group.

Because step 1 of the above algorithm does not involve communication between different CPU groups, very high parallel efficiency will be achieved for long MD sampling times, e.g., $T^* \geq 1$ ps. Since conformational transition rates of many macromolecules are typically low (1 μ s–1 ms), such sampling times are often required.

After the string has converged, the free energy profile is computed from additional MD simulations (which can be performed in parallel or in series) using Eqs. (33) and (34). An advantage of performing the simulations in parallel is that one can take advantage of the Hamiltonian replica exchange method (HREM) to accelerate convergence of the expectation values, as follows.⁷⁹

After a prescribed number of iterations, we attempt to switch the atomic coordinates of the MD replicas \hat{r}_i and \hat{r}_{i+1} that correspond to the CPU groups \mathcal{I} and $\mathcal{I} + 1$, respectively,

$$\begin{aligned} & \{\dots, (\hat{r}_i; \mathcal{I}), (\hat{r}_{i+1}; \mathcal{I} + 1), \dots\} \\ & \rightarrow \{\dots, (\hat{r}_{i+1}; \mathcal{I}), (\hat{r}_i; \mathcal{I} + 1), \dots\}, \end{aligned} \quad (45)$$

where i is chosen randomly from $0, 1, \dots, M - 1$. The acceptance probability of the move in Eq. (45) is computed according to the Metropolis criterion:

$$p_{acc}(\hat{r}_i \leftrightarrow \hat{r}_{i+1}) = \begin{cases} 1 & \text{if } \Delta \leq 0 \\ \exp(-\Delta) & \text{if } \Delta > 0, \end{cases} \quad (46)$$

where

$$\begin{aligned} \Delta = & \beta[U_k^{i+1}(\hat{r}_{i+1}) + U_k^i(\hat{r}_i) - U_k^{i+1}(\hat{r}_i) - U_k^i(\hat{r}_{i+1}) \\ & + V_\kappa^{i+1}(\hat{r}_{i+1}) + V_\kappa^i(\hat{r}_i) - V_\kappa^{i+1}(\hat{r}_i) - V_\kappa^i(\hat{r}_{i+1})]. \end{aligned} \quad (47)$$

The performance of the present implementation of the FTS is described in Sec. III for a β -sheet miniprotein, and in the supplementary material⁷⁴ for the alanine dipeptide.

III. RESULTS

In this section, we describe in detail the application of the rotation-invariant FTS to the β -sheet \leftrightarrow α -helix conformational transition of the 16-residue fragment of streptococcal protein G³⁷ (sequence 41-GEWYDDATKTFTVTE-56). This mini-protein has been used for testing the suitability of enhanced sampling methods to biomolecules.^{80–82} In addition,

the differences in the free energy, the enthalpy and the entropy between the two conformations were recently computed using a path-independent confinement method,⁸³ and are used here to validate the results of the present calculations. The results are organized as follows. In Sec. III A, we describe the generation of initial paths for the FTS. The application of the FTS to obtain converged transition paths is presented in Sec. III B. Profiles of the free energy computed by the restrained and tessellation methods are described in Sec. III C. We conclude the section with a description of the transition paths corresponding to the equilibrated paths in Sec. III D. The FTS was also applied to the alanine dipeptide (AD) in vacuum as a model system. The results for AD are presented in the supplementary material.⁷⁴

A. Initial minimum-energy paths

The preparation of the simulation structures is discussed in detail in Ref. 83, and is summarized only briefly here. A modified version of the program CHARMM^{77,84} version c37a1 was used in all calculations. The essential modifications involved implementing the restraint potentials U_k and V_κ , and an inter-processor communication scheme to collect string image coordinates necessary for reparametrization (see Sec. II). The Hamiltonian replica exchange method was based on the ENSEMBLE module in CHARMM. The modifications will be included in an upcoming release of CHARMM.

The β -sheet conformation was taken from the X-ray structure of protein G,³⁷ and the α -helical conformation was generated from default internal coordinate entries in CHARMM,⁸⁵ except for the backbone ϕ and ψ dihedrals, which were set to -57° and -47° , respectively. The all atom force-field with the CMAP correction⁸⁵ was used to represent the polypeptide, and the FACTS solvation model⁸⁶ was used to approximate the effects of solvent. The structures were subjected to 2000 steps of Adaptive-Basis Newton-Raphson (ABNR) minimization. The minimization of the α -helical structure was performed in the presence of harmonic restraints on the backbone dihedrals to maintain the α -helical conformation. The minimized structures were then equilibrated at 300 K in a 1 ns MD simulation with harmonic restraints on the backbone atom positions. An additional equilibration for 30 ns was performed without restraints. The α and β conformations after equilibration are shown in Fig. 1. The N-terminal turn of the α -helix conformation unwinds (Fig. 1(a)), but the rest of the helix is stable for the 30 ns MD simulation. In particular, the average and the standard deviation values of the helix backbone dihedral angles of res. 47–56 (i.e., excluding the partially unwound N-terminus) computed from the simulation were $(\phi, \psi) = (-63 \pm 7^\circ, -43 \pm 7^\circ)$. These values are within the rather broad α -helical region of the Ramachandran plot^{87–90} (for example, Best *et al.*⁹⁰ use the limits $[-30 < \phi < -100] \times [-67 < \psi < -7]$). We note that slightly different backbone dihedral values would result with the CHARMM36 force field, which includes a reoptimized backbone CMAP correction.⁹⁰

Transitions to the α form have been observed in previous MD simulations in implicit solvent starting from the β -sheet structure, and metastable states with significant helical

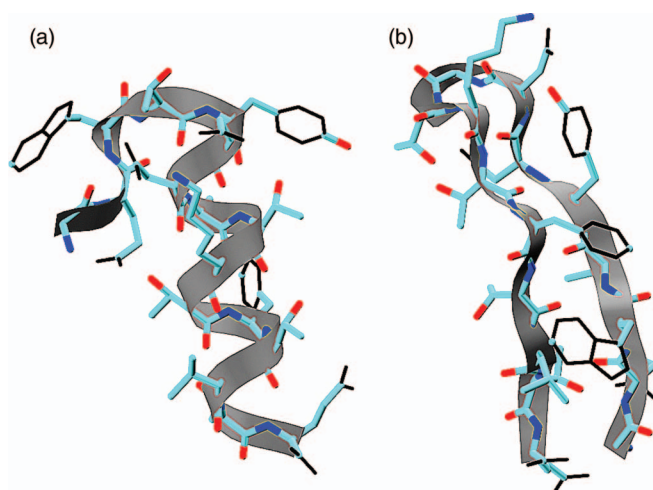


FIG. 1. Equilibrated structures of the 16-residue peptide 41-GEWTYDDATKTFVTVE-56 from protein G in (a) α -helical conformation, (b) β -sheet conformation. The N-terminal domain in panel (a) is at the top. Residue side chain atoms that are excluded from the definition of the string are drawn in black lines. Most of the excluded atoms are related by two-fold symmetry (i.e., C_D and C_E atoms in Phe and Tyr, terminal oxygens in Asp, Glu and in the C-terminal carboxyl group) and the remaining atoms belong to the aromatic rings of Trp (with the exception of C_H). Further details are given in the text.

content have been found in explicit solvent simulations (see review by Zhou⁹¹). Whether the α -helical conformation of the β -hairpin peptide actually plays physical role is unclear because solution experiments have not found evidence for appreciable α -helical content.⁹²

To initialize the FTS algorithm, two different MEPs were generated. The ZTS³⁶ method was used to generate the first path using six cycles of interpolation and minimization, as described previously.⁶⁷ In each cycle, a linear interpolation in Cartesian space that doubles the number of replicas was performed, followed by one thousand iterations of the ZTS method. Each iteration consists of 20 steps of steepest-descent (SD) minimization followed by a **reparametrization step**, to preserve equal distances between adjacent replicas. Because interpolation in Cartesian coordinates can lead to distortions in the local geometry (e.g., incorrect bond lengths, angles, or overlapping atoms), intermediate replicas may exhibit chirality errors. To maintain the correct (*L*-) amino-acid stereochemistry, the proton H_α on the C_α carbon was occasionally inverted through the plane normal to the $C_\alpha-H_\alpha$ bond. Similarly, to maintain the correct stereochemistry of the *Thr* C_β center, the proton H_β on the C_β carbon was occasionally inverted through the plane normal to the $C_\beta-H_\beta$ bond. To prevent the occurrence of cis-peptide bonds during minimization, weak dihedral-angle restraints centered on π (rad) with force constant $K_\omega = 2$ kcal/mol/rad² were added to each peptide bond (ω -dihedral angle). The destabilizing effects of chirality errors and cis-peptide bonds on the structure of small α -helical proteins in MD simulations were discussed recently by Schreiner *et al.*⁹³ The final 128-replica path was subjected to 25 000 additional ZTS iterations. The ZTS iterations are computationally inexpensive because they are dominated by SD minimization, and because each replica is assigned to a separate processor. The final ZTS minimization required

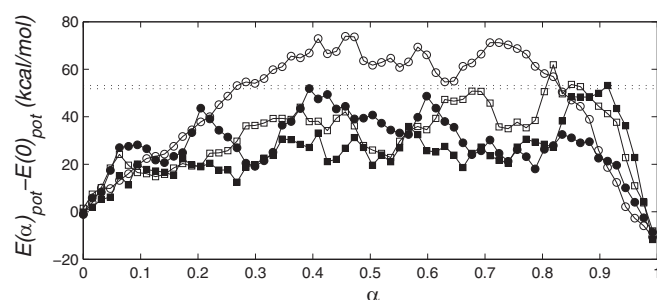


FIG. 2. Potential energy profiles along the MEPs that connect the α -helix and β -sheet conformations. Open symbols correspond to the MEPs from which the FTS iterations were initiated, and filled symbols correspond to the MEPs that were obtained by a minimization of the final equilibrated paths after the FTS converged (see text for details); —○— ZTS (initial); —●— ZTS (final); —□— RTMD (initial); —■— RTMD (final). The abscissa corresponds to the normalized string length ($\alpha = \int_0^\alpha d(\|\phi\|)/L$). The dotted horizontal lines are drawn at the vertical locations of the energy maxima of the paths after minimization.

approximately 1.5 h on 32 quad-core 2.4 GHz Opteron processors.

The second path was generated by a variant of **Targeted Molecular Dynamics**^{20,27,28} (TMD) using restraints,²⁹ as implemented in Ref. 30. A 2 ns Restrained TMD (RTMD) simulation was carried out, starting from the α -helical structure, and targeting the β -sheet structure. The initial RMSD between the two structures (≈ 9 Å) was reduced to the final RMSD value (which was below 1 Å) in Δ RMSD increments of 10^{-5} Å per timestep (2 fs, with constraints on bond lengths involving hydrogens⁹⁴). The target β -sheet structure was appended as the final point in the trajectory to ensure that the endpoints in the two initial paths were identical, and the resulting path was reduced to 128 replicas using linear interpolation. The reduced path was minimized using the ZTS method, as described for the final stage of the ZTS path above.

The potential energy profiles for the ZTS and RTMD paths are shown in Fig. 2 as a function of the normalized string length (α). The contributions to the potential energy for each MEP are listed in Table I (each value represents an average over the profile). From Fig. 2, we see that the initial MEP obtained from RTMD has significantly lower

TABLE I. Profile-average energy components for the MEPs in Fig. 2. The energy values corresponding to the first image ($\alpha = 0$) were subtracted from each average. The last row corresponds to the sum of the electrostatic energy and the polar solvation energy.

Energy contribution	ZTS (initial)	ZTS (final)	RTMD (initial)	RTMD (final)
Total	45.49	28.21	31.85	24.03
Bond	−0.41	0.16	−0.25	−0.30
Angle	2.14	0.98	1.70	1.52
Dihedral	11.55	13.25	13.88	12.87
Improper	−0.39	−0.45	−0.70	−0.42
Van der Waals	−0.89	−1.75	−1.12	−5.65
Electrostatic	94.08	55.81	10.73	1.53
Solvation (polar)	−57.20	−34.85	14.54	20.28
Solvation (nonpolar)	0.54	−1.06	−0.28	−0.87
Elec. + Solv.(p.)	36.88	20.95	25.28	21.81

energies (the maximum is around ~ 60 kcal/mol at $\alpha \simeq 0.82$) compared to the MEP obtained using ZTS (with a maximum of ~ 75 kcal/mol at $\alpha \simeq 0.45$). For this system, the TMD method appears to produce lower-energy transition paths than the ZTS. The contributions from the different energy terms^{77,86} listed in Table I differ between ZTS and RTMD mainly in the electrostatic and the polar solvation energies (computed using the FACTS model). In particular, the ZTS path has markedly higher electrostatic energy, which is offset by the lower polar solvation energy. The significance of the polar solvation term in the ZTS path suggests that the path proceeds through a more extensive exposure of the charged residues to the solvent, and the loss of favorable intramolecular interactions is compensated for by a gain in interactions with the solvent. To investigate this hypothesis quantitatively, we decomposed the polar solvation free energy into residue contributions, as implemented in the FACTS model in CHARMM, and compared the solvation energy profiles for the two paths for three different cases: all residues included, charged residues only (G41 [charged N-terminus], E42, D46, D47, K50, E56), and neutral residues only (43W, 44T, 45Y, 48A, 49T, 51T, 52F, 53T, 54V, 55T). The comparison is shown in Fig. S5a in the supplementary material.⁷⁴ If the total polar solvation energy includes only the neutral residues, the profile energies for the two paths have similar magnitudes. If only the charged residues are included, the profiles have similar differences as when all the residues are included. Examination of the motions of the charged residues along the transition paths were consistent with this analysis. For example, the number of hydrogen bonds (HBs) involving at least one sidechain atom of a charged residue follows a similar trend as the polar solvation energy profile (see Fig. S5b in the supplementary material⁷⁴), with the solvation energy being lower (i.e., more favorable) for replicas with fewer hydrogen bonds. (Following DeLoof, Nilsson, and Rigler,⁹⁵ who evaluated several HB identification criteria, we used the “COOR HBOND” facility in CHARMM with the simple criterion that the hydrogen-acceptor distance must not be greater than 2.4 Å.) Structures illustrating the differences in the hydrogen bonding involving charged residues are shown in Fig. S5, panels (c) and (d) of the supplementary material.⁷⁴

The differences in the residue motions that are responsible for the different electrostatic and polar solvation energies can be understood by considering the methods (RTMD and ZTS) used to obtain the initial paths. The RTMD path was obtained from a MD trajectory biased to evolve continuously from one endpoint to another along the potential energy landscape. Since the MD forces generally act in the direction opposite to the energy gradient vector (the thermostat produces additional small random forces), the simulated system resists overcoming energetic barriers, including the breaking of hydrogen bonds. In the ZTS method, the initial path is obtained by interpolation between then endpoint states, which does not take into account energies, possibly generating high-energy configurations. Although the subsequent minimization steps in the ZTS rapidly reduce high-energies due to distorted bonds, angles, torsions, and overlapping atoms, they generally do not restore broken hydrogen bonds or salt bridges if the separation between the groups involved is large. In such

cases, the electrostatic interactions are “screened” by the implicit solvation model, with a corresponding decrease in the polar solvation energy.

It is noteworthy that the average sum of the electrostatic and polar solvation energies has very similar values in the final ZTS and RTMD paths. As the above discussion suggests, the energetic compensation occurs because the electrostatic energy between protein atoms is computed in vacuum (i.e., with the dielectric of the medium set everywhere to unity), and is therefore overestimated. The FACTS solvation model essentially subtracts the overestimate and effectively shields the electrostatic interactions, as would be expected for a protein in solution. Profiles of the electrostatic energies, polar solvation energies, and their sum are shown in Fig. S6 of the supplementary material.⁷⁴ In addition, the difference between the energy maxima along the profiles corresponding to the annealed paths (i.e., *after* the application of FTS; see horizontal lines in Fig. 2) is only about 1.5 kcal/mol, even though the paths exhibit different sequences of events, and the maxima correspond to different α values. This fact underscores the importance of finite-temperature equilibration for comparing different MEPs.

Further discussion of the paths is deferred to Sec. III D, in which the transition mechanisms are described along with the corresponding free energy profiles.

B. Path optimization using FTS

Starting from the MEPs described in Sec. III A, equilibrated transition paths were computed using the FTS, as follows. To each of the 128 images we assigned a separate MD simulation system running on 4 processor cores, for a total of 512 cores per FTS simulation. The initial MD coordinates (as well as the initial string images) were taken from the MEPs computed above. Hydrogen atoms were not included in the definition of the string (equivalently, the weights w_i corresponding to the hydrogen atoms were zero). For reasons of computational efficiency, certain heavy side-chain atoms were also excluded from the string definition. These are (i) C_D and C_E carbon atoms in residues Phe and Tyr, (ii) terminal O_D and O_E oxygen atoms in Asp, Glu, and in the C-terminal carboxyl group, and (iii) atoms from the aromatic rings of Trp, with the exception of C_H . The excluded atoms are illustrated using black lines in Fig. 1. For convenience, the (N) atoms treated explicitly by the FTS are henceforth called *reaction atoms*, and the remaining atoms are called *bath atoms*. Reducing the number of reaction atoms decreases the number of rotation matrix derivatives needed to evaluate atomic forces and free energy derivatives, which must be calculated at each step of the MD calculation. In addition, removing ring and symmetric atoms prevents the path from lengthening artificially, e.g., if a ring were to flip during string equilibration. (Although ring rotation is often a rare event in equilibrium MD simulations, it is more likely to occur for high-energy intermediates along the string during evolution.)

The above choice of the reaction atoms assumes implicitly that the transition can be described accurately using only the positions of these atoms (after rigid body motions are removed). In principle, this assumption can be validated

by an *a posteriori* committor test.^{16,26,43,44} However, the full committor test can be impractical for diffusive or multi-barrier transitions, in which very long integration times are required for trajectories to commit to a basin. Using Markovian milestoning,^{67,75} we estimated the mean first passage times of the transition to be ≥ 7 ms (see Fig. S8 in the supplementary material⁷⁴). Given the ruggedness of the computed free energy profiles (discussed in Subsection III C), the full committor test would require integration times that are not affordable. We therefore performed a partial committor test for the dominant free energy barrier in the ZTS path, in which trajectories were only required to commit to the metastable states on either side of the barrier. Trajectories were launched from the two states along the reaction path adjacent to the highest free energy barrier. We found that 22% and 78% of the trajectories launched from the state closer to the product side terminate on the reactant and product sides, respectively, compared to 99% and 1% of the trajectories that were launched from the state closer to the reactant side. These values indicate that the transition state for the barrier (for which the “splitting” probability would be 50%/50%) is located between the two chosen states, as expected. The details of the calculation are given in Sec. SIV of the supplementary material.⁷⁴

Each replica was simulated in the canonical ensemble at 300 K using the Langevin dynamics thermostat coupled to heavy atoms with a friction constant of 1 ps^{-1} . Covalent bonds to hydrogen atoms were constrained with SHAKE,⁹⁴ and the FACTS implicit solvent model was used.⁸⁶ The same simulation conditions were used previously to calculate free energy differences by a confinement method.⁸³ The approximate time required to integrate 1 ns of MD was 2.85 CPU-core-hours per replica using a multiprocessor cluster equipped with quad-core AMD Opteron CPUs.

The parameters k , κ , and η_0 in the FTS restraint potentials U_k and V_κ (Eqs. (15) and (38), respectively) were selected empirically as follows. As done in previous studies,^{30,33,57,96} we employed a potential V_κ to limit the width of the transition tube around the string. The tube width parameter η_0 was set to 1 Å for all images, which was roughly equal to the RMS fluctuations computed from unrestrained endpoint MD trajectories. The tube width can also be estimated *a posteriori* (see Sec. IV). We note that the value of the tube width restraint can be important for the calculation of free energy profiles (discussed in Subsection III C), but should not have a significant effect on the converged path itself.

The restraint force constants k and κ were set to 10 kcal/mol/Å² and 20 kcal/mol/Å², respectively. The value of k is sufficiently high that the free energy profiles computed by the restrained method were in good agreement with those obtained by the tessellation method, yet low enough to achieve significant replica exchange acceptance frequencies during FE computation (discussed in Sec. III C). The value of κ was not optimized. However, unless κ is taken to be very low, its value is expected to have a much smaller effect on the free energy profile than that of the tube width parameter η_0 .

Starting from each MEP, the string was evolved from the initial condition until convergence, as follows. The running averages for the internal string images $\hat{\psi}$ were updated after each 10 successive MD steps, the averages were

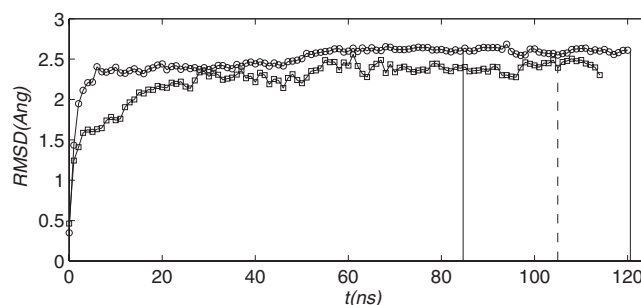


FIG. 3. Convergence of the FTS simulations monitored by the evolution of the distance ρ . —○— ZTS path evolution; —□— RTMD path evolution; the locations at which the string images were fixed for the calculation of the free profiles are marked by solid vertical lines for the ZTS path ($t_1 = 84$ ns and $t_2 = 121$ ns), and by a dashed line for the RTMD path ($t_1 = 107$ ns). The data are plotted in 1 ns increments, so that each interval corresponds to 500 reparametrizations.

reparametrized and the string images updated after 1000 MD steps. The string image update was performed gradually using linear interpolation over 100 MD steps (\mathcal{T}_{eq}) to avoid numerical instability. The running averages were not updated during this equilibration interval. The endpoint string images were not updated to facilitate comparisons between the free energy (FE) differences obtained from different paths (i.e., the endpoints are exactly the same).

Convergence of the paths was followed by monitoring the quantity $\rho(t)$ (Eq. (44)). The behavior of ρ is shown as function of simulation time in Fig. 3. Although convergence is reached after ~ 60 ns for each path, the simulations were continued to ~ 120 ns to ascertain that there was no systematic change in the evolution of ρ . For each simulation, the path RMSD between the initial and final paths was ~ 2.5 Å (which is relatively small compared to a RMSD of ~ 9 Å between the endpoint states), indicating that the string images do not travel very far from their initial positions, as was previously observed for simulations of myosin VI using the string method in collective variables.⁶⁷ (Further discussion is deferred until Sec. IV.) The calculation of FE profiles is discussed next.

C. Calculation of free energy profiles

FE profiles were computed for the instantaneous realizations of the strings marked by vertical lines in Fig. 3. To determine the sensitivity of the free energy profiles to dynamical fluctuations in the string coordinates within the converged region (e.g., after ~ 60 ns of simulation time), two time points corresponding to $t_1 = 84$ ns and $t_2 = 121$ ns were taken from the ZTS path. In principle, the time point at which the free energy profile is computed should not matter, as long as the corresponding string images are taken from the converged region, in which the fluctuations are small and random. We observed, however, that the computed profiles differ quantitatively, as described further in this subsection. For the RTMD path, a single realization of the string was taken as $t = 107$ ns.

The FE calculations performed with the tessellation (T) method, ZT1 and RT1 (Table II), do not have any adjustable parameters, and we take the corresponding profiles as a reference for testing the restrained (R) method, which requires

TABLE II. Summary of FE simulations performed in this study.

Index	Path	Method (R/T) ^a	t (ns)	k, κ, η_0	$\Delta F_{\alpha \rightarrow \beta}$ (kcal/mol)
ZT1 ^b	ZTS(t_2)	T	20	n/a	-8.25
ZR1	ZTS(t_2)	R	20	10, 20, 0.5	-8.4 ± 1.5
ZR2	ZTS(t_2)	R	20	10, 20, 1	-2 ± 3
ZR3	ZTS(t_2)	R	10	3, 20, 0.5	-5.6 ± 1.0
ZR4	ZTS(t_2)	R	20	10, 0, ∞	-1.6 ± 1.3
ZR5	ZTS(t_1)	R	15	10, 20, 0.5	-6.1 ± 1.8
RT1 ^b	RTMD	T	20	n/a	-8.3
RR1	RTMD	R	20	10, 20, 0.5	-8.2 ± 2.4
SCM					-6.7 ± 0.4

^aR and T refer to the *Restrained* and *Tessellation* methods for the calculation of FE profiles. SCM refers to the Simplified Confinement Method described in Ref. 83.

^bThe parameters k, κ , and η_0 are not relevant for the T method. In simulation ZR4, tube-width restraints were not used. ΔF is the free energy difference between the β -sheet and α -helix conformations.

values for the force constants and for the transition tube width. Although the tessellation method may be implemented with tube width constraints (see Sec. II) including such constraints did not make a significant difference for the present calculations.

The parameters of the FE simulations are summarized in Table II. The FE differences between the β and α conformations ($\Delta F_{\alpha \rightarrow \beta}$) computed from each profile are included. The value for $\Delta F_{\alpha \rightarrow \beta}$ computed by a path independent confinement method⁸³ is also listed. The restrained FE calculations were performed with the HREM to accelerate convergence, as described in Sec. II and discussed previously.⁶⁷ The exchange of adjacent MD replicas was attempted after 500 successive MD steps, and the average acceptance ratio was $\sim 40\%$ for the calculations (ZR1 and RR1) as discussed below.

The best match between the FE profiles computed using the T and R methods is obtained with the reference tube width η_0 set to 0.5 Å for both paths. The profile comparison is shown in Fig. 4. Although the profiles differ at multiple locations by several kcal/mol (e.g., $\alpha = 0.3$ –0.4 in panel (a)), the agreement is fair, given the differences in the sampling regions between the restrained and tessellation methods. (While the R method uses a uniform η_0 to sample the hyperplanes, in the T method the tube width is set implicitly to the sizes of the cells in the tessellation.) The profiles in Fig. 4 show similar $\Delta F_{\alpha \rightarrow \beta} \approx -8.3$ kcal/mol, although the uncertainties for the restrained method are significant, around $\simeq 2$ kcal/mol. Given the un-

certainty of the FTS methods, the value $\Delta F_{\alpha \rightarrow \beta} \approx -8.3$ is also within the error limits of the value $\Delta F_{\alpha \rightarrow \beta} \approx -6.7 \pm 0.4$ obtained by the (very different) SCM method.⁸³ The profiles corresponding to the remaining cases in Table II are discussed next.

The effect of the tube width η_0 on the FE profiles is illustrated in Fig. 5(a). In the absence of the restraint ($\eta_0 = \infty$), the FE profile fluctuates within ~ 3 kcal/mol around zero without displaying significant reaction barriers. This behavior is expected if the reaction coordinate isosurfaces are not very accurate, e.g., if they do not follow the actual FE barriers, as discussed by others.^{16,43} In the present calculation, some instantaneous configurations were more than 4 Å away from the string, where the hyperplane approximation is likely to be inaccurate. With $\eta_0 = 1$ Å (ZR2), the FE profile looks similar to that from simulation ZR1 ($\eta_0 = 0.5$). Because of the larger tube width, however, the convergence rate of the FE derivatives was slower, resulting in higher standard errors. This behavior was also observed in previous string calculations on the protein calmodulin,³⁰ in which a similarly narrow transition tube ($\eta = 0.75$ Å) was used to improve convergence of the FE profiles.

The effect of lowering the restraint force constant k on the FE profile is demonstrated in Fig. 5(b). As noted previously,⁶⁷ the effect of an insufficiently high force constant is to smooth the energy landscape by a convolution with a multivariate Gaussian with variances $(\beta k)^{-1}$. The FE profile corresponding to the force constant $k = 3$ (ZR3) has lower FE barriers, and a lower $\Delta F_{\alpha \rightarrow \beta}$, as expected from a smoothed landscape. We also note that the calculation ZR3 converged more rapidly than the other calculations because lower force constants increased the replica exchange frequency (to $\sim 70\%$), as observed previously.⁶⁷

Figure 5(b) also illustrates the effect of relatively minor variations in the string on the details of the FE profile. Simulations ZR1 and ZR5 use the instantaneous string obtained after 121 ns and 84 ns, respectively, and use identical simulation parameters. Both of the time instants are sampled after the RMS distance from the initial MEP ρ reaches a plateau (~ 60 ns in Fig. 3). However, small differences between the paths accumulate over the intervening 17 ns despite only small changes in $\rho(t)$, and result in differences in the FE profiles. The profiles are similar qualitatively, in that the locations of the FE minima and FE barriers are approximately preserved, but differ quantitatively. To understand the

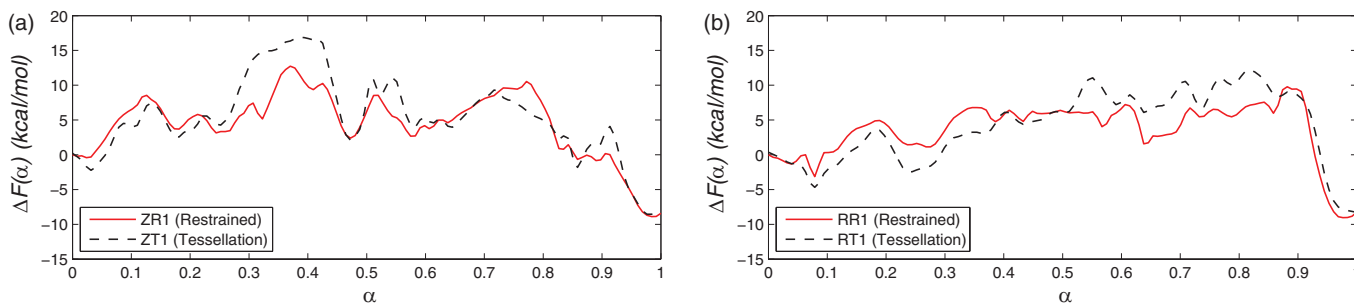


FIG. 4. Free energy profiles of transition ($\Delta F = F(\alpha) - F_0$) computed with the restrained (—) and tessellation (---) methods. (a) Initial path obtained using the linear interpolation combined with zero-temperature string (ZTS); (b) initial path obtained from a RTMD simulation and refined using ZTS.

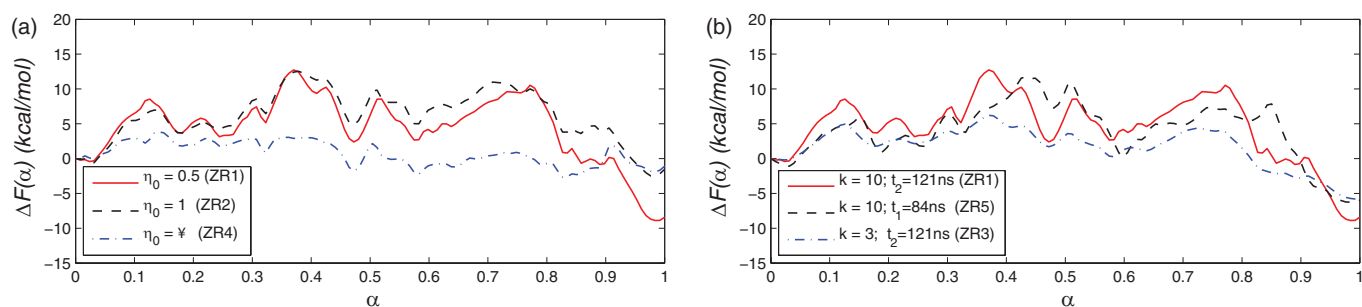


FIG. 5. (a) The influence of the transition tube width restraint (η_0) on the FE profile obtained using the restrained method. (b) Illustration of the effect of lowering the restraint force constant k (— vs. — · —), and of the dependence of the FE profile on the time point from which the path is taken (— vs. — · —). The initial MEP was generated using the ZTS method and linear interpolation.

source of the differences, we computed the RMS difference between the reaction atoms in the corresponding images for the two paths. The average distance was ~ 0.7 Å. The largest difference (~ 1.5 Å) was found near $\alpha \sim 0.8$. However, the difference in this region was below 1 Å if the images corresponding to 121 ns were shifted forward by ~ 0.07 (in terms of α). This suggests that, in the time interval $t = (84$ ns, 121 ns), the parametrization of the images at $\alpha \sim 0.8$ changed to $\alpha \sim 0.73$, i.e., the images were reindexed closer to the reactant basin, but did not undergo a significant structural change. This analysis (see also Sec. IV) reveals the limitations of the atom-based parametrization chosen for this study because of the high sensitivity of the string parametrization to small changes in the structure. This observation was an important motivation for the introduction of the string method in collective variables (SMCV),²⁶ developed for use with coarse-grained collective variables. Suitable coarse-grained variables, however, are not generally known for complex transitions, and for such cases (as for the case at hand), the present FTS method is the more appropriate choice, in spite of its limitations.

We conclude this section with a decomposition of the free energy profiles into the enthalpic and entropic contributions, shown in Fig. 6. The enthalpy profiles were computed by averaging the potential energy in the tessellation calculations ZT1 and RT2, and the entropic profile was obtained by subtracting from it the FE of the tessellation. While the enthalpy profile provides the larger contribution to the FE (15–20 kcal/mol), it is clear that the entropic “offset” is also significant, reaching values of ~ 10 kcal/mol at some locations. For example, in Fig. 6(a), the entropy term for

$\alpha = 0.7$ is ~ 8 kcal/mol higher than that of the α -helix, and ~ 13 kcal/mol higher than that of the β -sheet. It is noteworthy that the entropy of intermediate states tends to be higher than that of the endpoint states, suggesting the importance of multiple transition pathways. Furthermore, peaks in the entropy tend to accompany enthalpic peaks, thereby reducing the overall free energy barrier. This observation suggests the importance of enthalpy-entropy compensation in protein conformational transitions, which has been observed experimentally in protein-ligand association.^{97–99} Specifically, the binding of a ligand lowers the enthalpy of the bound complex, compared to the unbound form, but does so at the expense of (1) removing the entropy in the rotational and translational degrees of freedom (DOFs) of the ligand that are eliminated in the binding (the internal DOFs created by the formation of the complex contain significantly less entropy because the ligand is effectively immobilized in the protein binding site^{100,101}) and (2) reducing the conformational entropy of the protein residues involved in the binding, whose mobility is lower due to the favorable enthalpy of binding. In the case of conformational transitions, the protein must proceed through intermediate (higher-energy) states to switch between (low-energy) endpoint states; when the number of possible intermediate configurations is large, compared to those for the endpoints, the enthalpic barriers will be partially offset by entropic effects, as observed in Fig. 6.

D. Illustration of transition paths

To illustrate the transition, eight string images were selected from each path. The coordinates of the reaction atoms

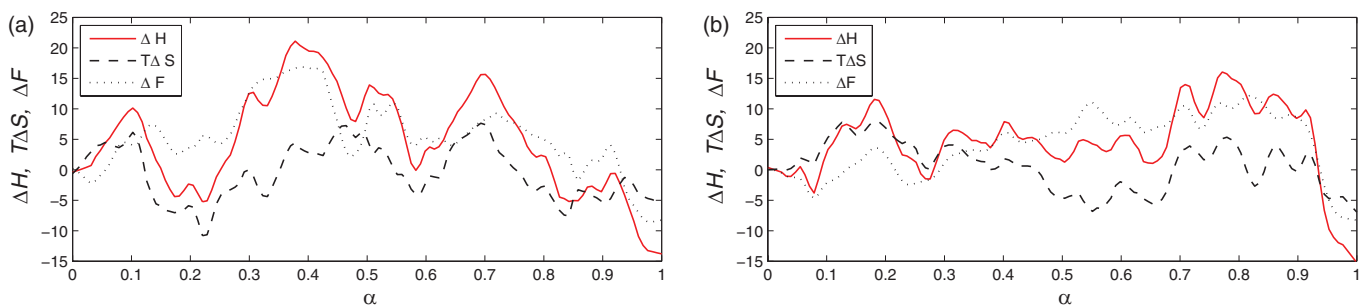


FIG. 6. Decomposition of the FE profiles computed with the tessellation method (· · · · ·) into enthalpic (—) and entropic (— — —) contributions. (a) Simulation ZT1; (b) simulation RT1.

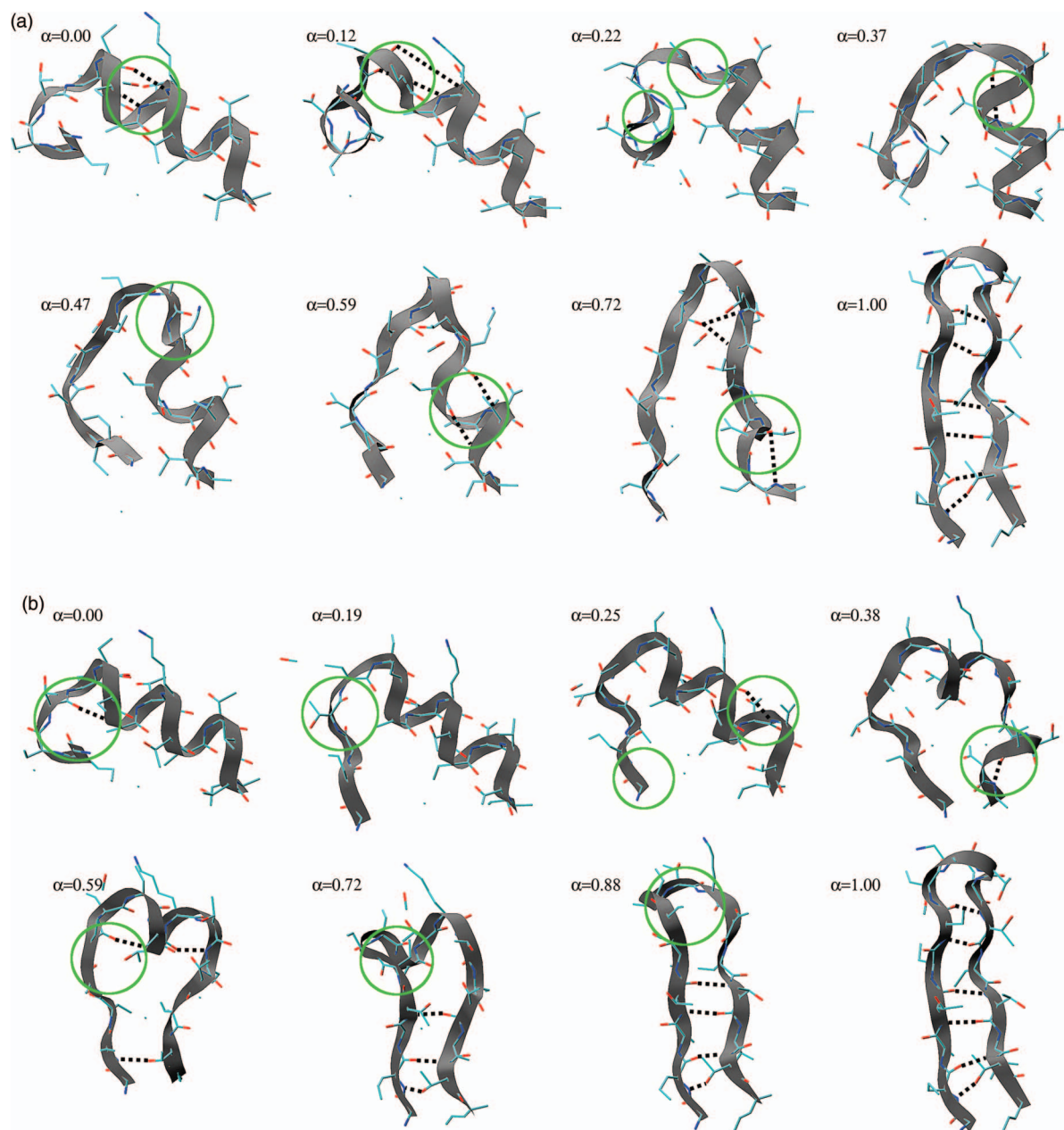


FIG. 7. Snapshots from finite temperature paths that illustrate the transition mechanism. (a) Initial path obtained using the linear interpolation combined ZTS, $t = 121$ ns (b) initial path obtained from a RTMD simulation and refined with ZTS, $t = 107$ ns. The peptide backbone is drawn in ribbon representation, and the reaction atoms are drawn in ball-and-stick representation. Forming or breaking hydrogen bonds are indicated by dotted lines. For each step, regions of the protein that are undergoing significant structural changes are highlighted with a green circle. Bath atoms are not shown because they are not part of the string definition (their string coordinates are undefined). All of the atoms are shown in Fig. 1. (Multimedia view) [URL: <http://dx.doi.org/10.1063/1.4871685.1>] [URL: <http://dx.doi.org/10.1063/1.4871685.2>].

are shown in the $\alpha \rightarrow \beta$ direction in Figures 7(a) (Multimedia view) and 7(b) (Multimedia view) for the ZTS and RTMD paths, respectively. The converged paths differ quantitatively in terms of the associated FE profiles (see Fig. 4), but also qualitatively in terms of the sequence of atomic motions. The first FE barrier along the ZTS path occurs for $\alpha \simeq 0.12$ (Fig. 4(a)), corresponding to the unwinding of the first helical loop at the N-terminal end, which requires the breaking of intra-helical hydrogen bonds. The next FE barrier at $\alpha \simeq 0.22$ corresponds to twisting motions within the partially unwound loop, which results in a nearly straight β -strand ($\alpha \simeq 0.37$). The dominant FE barrier of the ZTS transition path (ΔFE

$\simeq 14$ kcal/mol) corresponds to the formation of the hairpin (images at $\alpha \simeq 0.37$ and $\alpha \simeq 0.47$). It is noteworthy that this event also corresponds to the highest barrier in the RTMD path, although in that path it occurs at a much later stage ($\alpha \simeq 0.88$ in Fig. 4(b)). The images corresponding to $\alpha \simeq 0.47$ and $\alpha \simeq 0.59$ are the most significant metastable states along the ZTS-derived path, with the FE profile indicating barriers of ~ 7 kcal/mol on either side of the minima. The images correspond to “hybrid” structures in which the N-terminal half of the protein is a β -strand, and the C-terminal end has 1–2 turns of the α -helix. The final barrier corresponds to the unwinding of the remaining α -helical turn at $\alpha \simeq 0.72$, which is

followed by intra-strand hydrogen bond formation that leads to the β -sheet.

The sequence of transition events along the RTMD path is different, and is consistent with previous analyses of paths generated by TMD,^{28,102–104} in which large-scale rearrangements precede small-scale motions.³⁰ In particular, the initial stage of transition is dominated by a large displacement of the N-terminal end toward the C-terminus ($\alpha \simeq 0$ to $\alpha \simeq 0.38$), and the final stage ($\alpha \geq 0.88$) consists of local adjustments to the conformation of the hairpin head. The transition begins with the unwinding of the first N-terminal helical loop ($\alpha \simeq 0.19$) in a similar fashion and with a similar FE barrier as observed in the ZTS path, except that in this case the loop rapidly moves toward its final position in the β -hairpin. The C-terminal helical turn unwinds next, and the N-terminus moves further toward the C-terminus to form the initial hydrogen bond between the strands of the emerging hairpin ($\alpha \simeq 0.25$ to $\alpha \simeq 0.6$). These events are accompanied by a gradual increase in the FE along the profile (Fig. 4(b)) with rather shallow metastable states (barriers of 5 kcal/mol or less). The most metastable state is structurally similar to the image at $\alpha \simeq 0.59$, and shows a hybrid structure, with the interior of the peptide sequence α -helical, and the ends resembling a β -sheet. The dominant FE barrier along this path occurs near $\alpha \simeq 0.88$, and corresponds to the formation of the correct hairpin.

IV. DISCUSSION

Calculating free energy profiles for conformational changes in proteins from molecular simulations is challenging because an appropriate reaction coordinate is difficult to determine. In the present study, we used the finite temperature string method (FTS) in rotation-invariant coordinates to approximate the reaction coordinate isosurfaces for a 16-residue β -hairpin. Free energy profiles were computed either by sampling the hyperplanes directly using restrained dynamics, or by constructing tessellations of the configurational space associated with the string. A central assumption in the FTS is that the most probable transition paths connecting reactants and products are clustered together in relatively narrow transition tubes in configurational space. Conceptually, this assumption implies that all transition events occur in a similar fashion, and therefore can be represented in an average sense, i.e., by a “string” of images that correspond to typical configurations at each stage of the reaction. Mathematically, the narrow-tube assumption is used to justify approximating the reaction coordinate isosurfaces by hyperplanes in the vicinity of a maximum in the density of reactive trajectories.^{25,26,40,67} This approximation is crucial as it makes sampling the reaction coordinate isosurfaces computationally tractable. Qualitative support for the assumption can be found in the micro- to millisecond long MD trajectories obtained by Shaw and co-workers, who observed that most folding events for the WW domain Fip35 occurred essentially in the same order, and that the conformational transitions in the protein BPTI proceed *via* a small number of structurally distinct states.^{105,106} An independent analysis of the Fip35 folding trajectories¹⁰⁵ using Markov state modeling (MSM) concluded that there are

multiple pathways to the native structure,¹⁰⁷ in accord with earlier MSM predictions based on short MD simulations.¹⁰⁸ Nevertheless, the folding pathway observed by Shaw *et al.*¹⁰⁵ was found to carry the majority of the reactive flux, indicating that the other pathways are less important kinetically. Generally, the accuracy of the narrow tube assumption in the FTS method will depend on the specific system under study; the FTS method employed here will be most useful for transitions for which the narrow-tube assumption is accurate.

Quantitatively, the transition tube width is difficult to estimate because it depends on the coordinates chosen to represent the actual transition path ensemble (here, the string). In the present calculations, we used restraint potentials to investigate the effect of an imposed tube width on the FE profile, and found a strong dependence (Fig. 5(a)). To obtain agreement between the FE profiles computed by tessellation and those computed using restraints, it was necessary to limit MD sampling in the restrained calculations to relatively narrow tubes of radius $\simeq 0.5$ Å around the string. This value also gives the approximate range of validity of the hyperplane approximation (in the present atom-based coordinates). In the absence of tube width restraints, the restrained method produced an unrealistic FE profile with small barriers (Fig. 5(a)), as would be expected for an incorrect reaction coordinate.^{16,43} Furthermore, using the instantaneous configurations from MD sampling of the tessellation in simulations ZT1 and RT1 (Table II) the distances from the string to the MD replicas were found to vary from 0.5 Å to 2.5 Å along the transition path, roughly consistent with the value 0.5 Å used in the restrained simulations. The strong dependence of the FE profiles on the imposed tube width indicates that the narrow-tube assumption is inaccurate for some transitions (and/or some variables) and requires additional tube restraints to avoid sampling incorrect regions. Because the tessellation method prevents the crossing of hyperplanes by construction, it appears to be technically superior to the restraint method, which requires (*ad hoc*) specification of tube restraints. However, because neither method can guarantee the validity of the hyperplane approximation away from the string, comparing the FE profiles obtained from both methods, as done here, is useful for qualitative validation.

The requirement of a narrow tube restraint found in this study reflects, in part, the difficulty of finding suitable coordinates for coarse-graining the energy landscape. In a previous study, we considered several types of collective variables (CVs) to represent the string for a transition in the converter domain of Myosin VI.⁶⁷ In that case, however, the choice of CVs was simplified because the differences in the endpoint conformations could be described by reorientations of α -helices and a β -sheet, connected by short loops. More generally, for transitions that involve complex motions of many atoms, such as the protein secondary structure changes considered here, finding appropriate coarse-grained CVs can be very difficult.

The choice of coordinates to represent the string also determines the dimensionality of the FE landscape. A larger number of coordinates will generally produce a more rugged landscape with more possible transition paths, which also

require a finer discretization. In the present case, the FTS simulations were initialized from two different MEPS, each exhibiting a different order of conformational changes. Each FTS simulation converged in the vicinity of the initial path (RMS distance of ~ 2.5 Å, see Fig. 3). The two very distinct paths had similar free energy profiles, indicating that both paths should be taken into account in describing the ensemble of actual transition paths. Because the free energy profile, *per se*, does not include information on the reaction rate, we estimated the mean first passage times (MFPT) from the α -helical state to the β -sheet state along the two paths using Markovian milestoneing.⁹⁵ The MFPTs were found to be $\simeq 7$ ms and $\simeq 200$ ms for the RTMD and ZTS paths, respectively, suggesting that the RTMD path carries more reactive flux. This result is also consistent with the fact that the computed FE barrier for the RTMD path is slightly lower than that for the ZTS path, although the difference must be interpreted in light of the ~ 2 kcal/mol uncertainty in the FE profile values. The existence of the two distinct transition paths described here suggests that a larger number of paths (in the present string coordinates) could be required to characterize the transition completely. Alternatively/additionally there may exist paths not found by the present approach, along which the free energy barriers are lower, which carry more reactive flux, and which would be more kinetically significant than the paths described here. Clearly, the need to account for multiple paths remains a challenge in the study of protein folding and transitions by all presently available methods, including the FTS.

An important conclusion from the enthalpy/entropy decomposition of the FE profiles (Fig. 6) is that entropic differences can offset the enthalpic barriers in conformational transitions by as much as 10 kcal/mol. This implies that in the vicinity of energy barriers, the actual transition trajectories can follow a large number of paths. For the present β -hairpin system this is consistent with partially unfolded configurations (at e.g., $\alpha \simeq 0.45$ or $\alpha \simeq 0.7$ in Figs. 6(a) and 7(a)). In addition, the contributions from the entropy may actually be underestimated because of the limits on the transition tube width (which are implicit in the tessellation calculation).

A related point concerns the differences between minimum energies and free energies computed along each transition tube. A comparison between the profiles in Fig. 6 and the minimum energy profiles annealed from the string images in Fig. 2 shows that the values along the minimum energy profiles can be 20–30 kcal/mol higher than those of the free energy. This further highlights the fact that regions of configurational space large enough to capture entropic contributions to the free energy, rather than single-point configurations, should be considered when evaluating the stability of a particular state.

Finally, to the extent that the rigid-body motions are separable from the internal configurational motions, the present FTS algorithm computes only the internal part of the free energy, since rigid-body motions are not restrained. While the translational contribution to the total partition function is a constant that is proportional to the concentration, the rotational contribution will be variable along the string, because it is proportional to the square root of the inertia tensor determinant, so that an extended system is entropically

favored over a compact system. To estimate the rotational contribution to the total free energy, we computed the inertial tensor determinants for the string corresponding to simulation ZTS(t_2) (see the supplementary material⁷⁴). The β -sheet conformation is favored over the α -helical conformation by about 0.1 kcal/mol, and the intermediate configurations around $\alpha = 0.25$ are disfavored relative to the α -helix by about 0.12 kcal/mol. These trends are qualitatively consistent with the snapshots in Fig. 7(a), in which the β -sheet configurations are the most extended, and the transitional structures around $\alpha = 0.22$ are the most compact. The magnitude of the rotational corrections, however, is about an order of magnitude less than the standard error in the configurational free energy, and the corrections are therefore not significant for the present system.

V. CONCLUSION

We described a parallel implementation of the FTS method in rotation-invariant and translation-invariant coordinates that use the degenerate distance metric of Vandeneijnden and Venturoli⁴⁶ with hypersurface restraints to compute one-dimensional free energy profiles. The main difference between the present approach and earlier FTS applications to rigid-body invariant systems,^{50,51} is that coordinates are implicitly transformed to the reference frames of the corresponding string images, eliminating the need for arbitrarily predefined reference structures. The method was applied to the β -sheet $\leftrightarrow\alpha$ -helix transition in a 16-residue miniprotein to calculate transition tubes and free energy profiles associated with two distinct initial paths. The computed free energy differences were found to be within the uncertainty limits of the value obtained previously from a confinement free energy analysis.⁸³ The present results demonstrate the utility of the FTS for computing transition paths and the associated free energy profiles in complex biomolecular systems. In particular, the transition considered here involves a complete change of the protein secondary structure, for which it is difficult to find a reaction coordinate,^{43–45} or even a suitable set of collective variables. The presence of many reaction barriers in the free energy profiles (e.g., Fig. 4) would also make it nontrivial to use methods such as TPS,^{15,16} as each barrier would require a separate application.¹⁰⁹ Thus, the FTS appears to be well-suited for obtaining transition paths and free energies in complex systems such as the one examined here. A technical caveat associated with the calculation of the free energy profiles concerns the accuracy of the hyperplane approximation implicit in the FTS. In particular, FE profiles computed using a method based on harmonic restraints to the hyperplane require the additional specification of a tube width restraint, whose reference value is generally determined by trial and error. The dependence of the computed free energy profiles on the tube width implies that the hyperplane approximation is not very accurate in regions farther than ~ 1 Å away from the string, using the atom-based coordinates employed here. A possible direction for the construction of coarser-grained coordinates better-suited for generic protein transitions could start from bead models, in the spirit of the moderately coarse-grained residue

representation used in the MARTINI force field,¹¹⁰ with multiple interaction sites per residue, and proceed systematically toward further coarse-graining, using, e.g., a single interaction site model.^{111,112} The use of such coarse graining schemes would require only minor modifications to the FTS implementation outlined here. At each stage of coarse-graining, the FE profile computed based on the reduced set of coordinates can be validated with respect to those computed using finer sets. These explorations are left to a future study.

ACKNOWLEDGMENTS

The authors are grateful to Eric Vanden-Eijnden for a critical reading of the paper and for suggestions that improved the final version. V.O. is grateful to Arnaud Blondel for a discussion of amino acid stereochemistry in transition path optimization. Computer resources were provided by the National Energy Resource Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the Faculty of Arts and Sciences Computing Group at Harvard. The work done at Harvard was supported in part by the National Institutes of Health.

APPENDIX: CALCULATION OF THE METRIC TENSOR AND PROJECTION DERIVATIVES

First, we compute the metric tensor $M_{i,j}^{a,b}$ defined by Eq. (6). Noting that the derivatives with respect to the absolute coordinates (\hat{r}^*) are related to those with respect to the coordinates relative to the COM by $\partial/\partial r_k^{*c} = \partial/\partial r_k^c - w_k \sum_{i=1}^N \partial/\partial r_i^c$, and using the identity³⁰ $\sum_{i=1}^N \partial A/\partial r_i^c = 0$, we obtain

$$M_{i,j}^{a,b} = \sum_{c=1}^3 \sum_{k=1}^N \frac{1}{m_k} \left\{ \left\{ \frac{\partial A^{ad}}{\partial r_k^c} r_i^d + A^{ac}(\delta_{ik} - w_k) \right\} \times \left\{ \frac{\partial A^{be}}{\partial r_k^c} r_j^e + A^{bc}(\delta_{jk} - w_k) \right\} \right\} = \frac{1}{m_j} \delta_{ab} \delta_{ij} - \delta_{ab} \left(\sum_{k=1}^N m_k \right)^{-1} + \sum_{c=1}^3 \sum_{k=1}^N \frac{1}{m_k} \left\langle r_i^d \frac{\partial A^{ad}}{\partial r_k^c} \frac{\partial A^{be}}{\partial r_k^c} r_j^e \right\rangle + \sum_{c=1}^3 \left\langle \frac{1}{m_j} r_i^d \frac{\partial A^{ad}}{\partial r_j^c} A^{bc} + \frac{1}{m_i} r_j^e \frac{\partial A^{be}}{\partial r_i^c} A^{ac} \right\rangle, \quad (\text{A1})$$

where summation over the indices d and e is implied, and the transposition operator is omitted from A for brevity. The first term on the right-hand side of Eq. (A1) is the inverse mass tensor, and the remaining terms constitute the difference tensor $D_{i,j}^{a,b}$ (defined in Eq. (7)). Figure 8 shows the normalized tensor $|D_{i,j}^{a,b}|(m_i m_j)^{1/2}$ computed for the β -hairpin molecule, averaged over the 128 replicas in the converged path ZTS(t_2) (see Table II). The average deviation from zero is $\simeq 0.005$, and the highest deviation is $\simeq 0.045$. The largest contributions to the deviations arise from the last term in Eq. (A1). Because the term $\delta_{ab}(\sum_{k=1}^N m_k)^{-1}$ contributes only to the diagonals of the 3×3 coordinate submatrices, the tensor has a “grainy” ap-

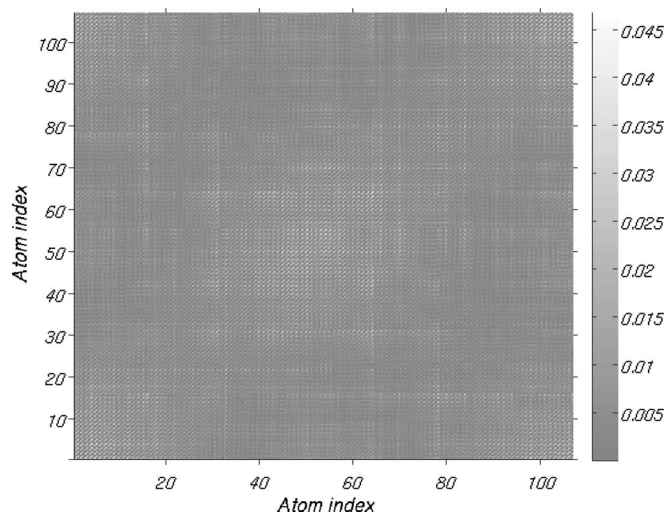


FIG. 8. Normalized deviation tensor $|D_{i,j}^{a,b}|(m_i m_j)^{1/2}$ for the β -hairpin mini protein (see Eq. (A1)). The filled contours are shown as a series of adjacent 3×3 submatrices, i.e., the component indices a and b vary faster than the coordinate indices i and j in the horizontal and vertical directions, respectively.

pearance. $|D_{i,j}^{a,b}|(m_i m_j)^{1/2}$ for the alanine dipeptide is shown in the supplementary material.⁷⁴

For a regular coordinate transformation, the Jacobian can be computed from the (non-mass-weighted) metric tensor using $J = (\det M)^{1/2}$. However, the transformation $\hat{r}^* \mapsto A^T \hat{r}$ is not invertible, because any rotation and/or translation of the absolute coordinates \hat{r}^* maps to the same (invariant) coordinates $A^T \hat{r}$, and M in this case is singular with a zero determinant. The degeneracy of M can be removed by computing the “configurational determinant,” i.e., omitting the six lowest eigenvalues of M , which correspond to rotation and translation and are nearly zero, with the remaining eigenvalues close to unity. This procedure is analogous to normal mode calculations, in which the six lowest-frequency normal modes are excluded from analysis because they correspond to rigid motions.¹¹³ The Jacobian is then calculated as

$$J \equiv |\nabla_{\hat{r}^*}(A^T \hat{r})| = \left(\prod_{i=7}^{3N} \lambda_i \right)^{1/2}, \quad (\text{A2})$$

where the λ_i are the eigenvalues of M computed with the masses set to unity to avoid mass-weighting, and N is the number of string atoms. Figure 9 shows the variation of $J(\alpha)$ along the string. Both, an instantaneous curve and a curve averaged over the 20 ns ZT1 trajectory (see Table II) are plotted. The Jacobian fluctuates within about 3% of the average value, and it is expected that the fluctuations would make negligible contributions to the free energy profile. For example, $k_B T \log 1.03 \approx 0.017$ kcal/mol, which is well below the standard errors reported in Table II. Therefore, the omission of the Jacobian factor from Eq. (11) is fully justified.

Next, we compute the gradients of the quantities ϵ_i and η_i that are required to apply the forces arising from the potentials $U_k(\hat{r})$ and $V_k(\hat{r})$. Each ϵ_i in Eq. (27) can be rewritten more concisely as

$$\epsilon = [\hat{r} - A\hat{p}] \cdot [B\hat{q} - A\hat{p}]/\Delta, \quad (\text{A3})$$

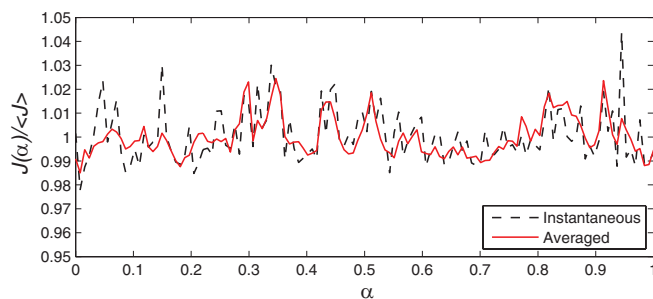


FIG. 9. Jacobian of the transformation $\hat{\mathbf{r}}^* \mapsto A^T \hat{\mathbf{r}}$ normalized by its average over the string in simulation ZT1: (---) instantaneous curve computed at the beginning of the simulation; (—) simulation average.

where $\hat{\phi}$ and \hat{q} are constant, $A(\hat{\mathbf{r}}, \hat{p})$ and $B(\hat{\mathbf{r}}, \hat{q})$ minimize the norms $\|\hat{\mathbf{r}} - A\hat{p}\|$ and $\|\hat{\mathbf{r}} - B\hat{q}\|$, respectively, and $\Delta = \|B\hat{q} - A\hat{p}\|$. Using (\cdot) to represent partial derivatives with respect to r_i^t (t -coordinate of atom i), we have

$$\Delta' = -\frac{1}{\Delta} \sum_{k=1}^N w_k \hat{p}_k^T (A^T B)' \hat{q}_k^T \quad (\text{A4})$$

and

$$\epsilon' = \frac{1}{\Delta} \{w_i (B\hat{q}_i)^t - w_i (A\hat{p}_i)^t\} + (1 - \epsilon/\Delta) \Delta'. \quad (\text{A5})$$

The computation of derivatives of rotation matrices is described in the supplementary section of Ref. 30. Derivatives of $\eta = \|v_\perp\|$ can be computed with very little additional effort by noting that (with A and B as in Eq. (A3))

$$v^\perp = (1 - \epsilon/\Delta)(\hat{\mathbf{r}} - A\hat{p}) + (\epsilon/\Delta)(\hat{\mathbf{r}} - B\hat{q}) \quad (\text{A6})$$

(compare with Eq. (39)), which implies

$$[v^\perp]' \cdot [v^\perp] = w_i (v^\perp)_i^t - \epsilon(1 - \epsilon/\Delta) \Delta'. \quad (\text{A7})$$

We note that in each of Eqs. (A5) and (A7), the last term on the right-hand side is due to rotation matrix derivatives. The additional terms are needed to keep the MD replica on the reaction coordinate hypersurfaces because the optimal rotations will change when the coordinates are perturbed.

¹S. Benkovic and S. Hammes-Schiffer, *Science* **301**, 1196 (2003).

²K. Johnson, *Annu. Rev. Biochem.* **62**, 685 (1993).

³C. Joyce and T. Steitz, *Annu. Rev. Biochem.* **63**, 777 (1994).

⁴J. Ma, P. Siegler, Z. Xu, and M. Karplus, *J. Mol. Biol.* **302**, 303 (2000).

⁵M. Karplus and Y. Gao, *Curr. Opin. Struct. Biol.* **14**, 250 (2004).

⁶A. Houdusse and H. L. Sweeney, *Curr. Opin. Struct. Biol.* **11**, 182 (2001).

⁷M. Geeves and K. Holmes, *Annu. Rev. Biochem.* **68**, 687 (1999).

⁸E. A. Morrison, G. T. DeKoster, S. Dutta, R. Vafabakhsh, M. Clarkson, A. Bahl, D. K. T. Ha, and K. Henzler-Wildman, *Nature (London)* **481**, 45 (2012).

⁹R. Dror, D. Arlow, D. Borhani, M. Jensen, S. Piana, and D. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4689 (2009).

¹⁰M. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A. Beberg, D. Ensign, C. Bruns, and V. Pande, *J. Comput. Chem.* **30**, 864 (2009).

¹¹M. Harvey, G. Giupponi, and G. D. Fabritis, *J. Chem. Theory Comput.* **5**, 1632 (2009).

¹²G. Torrie and J. Valleau, *J. Comput. Phys.* **23**, 187 (1977).

¹³C. Bartels, M. Schaefer, and M. Karplus, *J. Chem. Phys.* **111**, 8048 (1999).

¹⁴A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).

¹⁵C. Dellago, P. Bolhuis, and P. Geissler, *Adv. Chem. Phys.* **123**, 1 (2003).

¹⁶P. Bolhuis, D. Chandler, C. Dellago, and P. Geissler, *Ann. Rev. Phys. Chem.* **53**, 291 (2002).

¹⁷J. Hénin and C. Chipot, *J. Chem. Phys.* **121**, 2904 (2004).

¹⁸J. Hénin, G. Fiorin, C. Chipot, and M. L. Klein, *J. Chem. Theory Comput.* **6**, 35 (2010).

¹⁹L. Rosso and M. Tuckerman, *Mol. Sim.* **28**, 91 (2002).

²⁰J. Schlitter, M. Engels, P. Krüger, E. Jacoby, and A. Wollmer, *Mol. Sim.* **10**, 291 (1993).

²¹L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.* **426**, 168 (2006).

²²A. Faradjian and R. Elber, *J. Chem. Phys.* **120**, 10880 (2004).

²³D. Shalloway and K. Faradjian, *J. Chem. Phys.* **124**, 054112 (2006).

²⁴S. Kirmizialtin, V. Nguyen, K. Johnson, and R. Elber, *Structure* **20**, 618 (2012).

²⁵W. E, W. Ren, and E. Vanden-Eijnden, *Chem. Phys. Lett.* **413**, 242 (2005).

²⁶L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *J. Chem. Phys.* **125**, 024106 (2006).

²⁷J. Schlitter, M. Engels, and P. Krüger, *J. Mol. Graph.* **12**, 84 (1994).

²⁸A. van der Vaart and M. Karplus, *J. Chem. Phys.* **122**, 114903 (2005).

²⁹J. Apostolakis, P. Ferrara, and A. Caffisch, *J. Chem. Phys.* **110**, 2099 (1999).

³⁰V. Ovchinnikov and M. Karplus, *J. Phys. Chem. B* **116**, 8584 (2012).

³¹T. Schlick, *F1000 Biology Reports* **1**, 51 (2009).

³²H. Lei and Y. Duan, *Curr. Opin. Struct. Biol.* **17**, 187 (2007).

³³W. Ren, E. Vanden-Eijnden, P. Maragakis, and W. E, *J. Chem. Phys.* **123**, 134109 (2005).

³⁴G. Jónsson, and K. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. Berne, G. Ciccotti, and D. Coker (World Scientific, Singapore, 1998), pp. 385–404.

³⁵S. Fischer and M. Karplus, *Chem. Phys. Lett.* **194**, 252 (1992).

³⁶W. E, W. Ren, and E. Vanden-Eijnden, *J. Chem. Phys.* **126**, 164103 (2007).

³⁷A. Gronenborn, D. Filpula, N. Essig, A. Achari, M. Whitlow, P. Wingfield, and G. Clore, *Science* **253**, 657 (1991).

³⁸W. E and E. Vanden-Eijnden, *Annu. Rev. Phys. Chem.* **61**, 391 (2010).

³⁹W. E and E. Vanden-Eijnden, in *Multiscale Modeling and Simulation*, edited by S. Attinger and P. Koumoutsakos (Springer-Verlag, Berlin, 2004).

⁴⁰W. E, W. Ren, and E. Vanden-Eijnden, *Phys. Rev. B* **66**, 052301 (2002).

⁴¹C. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. (Springer-Verlag, Berlin, 2003).

⁴²G. Hummer, *J. Chem. Phys.* **120**, 516 (2004).

⁴³R. Best and G. Hummer, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6732 (2005).

⁴⁴A. Ma and A. Dinner, *J. Phys. Chem. B* **109**, 6769 (2005).

⁴⁵B. Peters and B. Trout, *J. Chem. Phys.* **125**, 054108 (2006).

⁴⁶E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **130**, 194103 (2009).

⁴⁷A. Dickson, A. Warmflash, and A. R. Dinner, *J. Chem. Phys.* **130**, 074104 (2009).

⁴⁸E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **131**, 044120 (2009).

⁴⁹A. Dickson, A. Warmflash, and A. R. Dinner, *J. Chem. Phys.* **131**, 154104 (2009).

⁵⁰I. Khavrutskii, K. Arora, and C. Brooks III, *J. Phys. Chem.* **125**, 174108 (2006).

⁵¹A. Pan, D. Sezer, and B. Roux, *J. Phys. Chem. B* **112**, 3432 (2008).

⁵²J. L. Farrell, J. C. Stuelpnagel, R. H. Wessner, J. R. Velman, and J. E. Brook, *SIAM Rev.* **8**, 384 (1966).

⁵³W. Kabsch, *Acta Cryst.* **A32**, 922 (1976).

⁵⁴B. Horn, *J. Opt. Soc. Am. A* **4**, 629 (1987).

⁵⁵E. Coutsias, C. Seok, and K. Dill, *J. Comput. Chem.* **25**, 1849 (2004).

⁵⁶R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987).

⁵⁷E. Neria, S. Fischer, and M. Karplus, *J. Chem. Phys.* **105**, 1902 (1996).

⁵⁸J.-W. Chu, B. Trout, and B. Brooks, *J. Chem. Phys.* **119**, 12708 (2003).

⁵⁹R. Czereminski and R. Elber, *J. Chem. Phys.* **92**, 5580 (1990).

⁶⁰C. Eckart, *Phys. Rev.* **47**, 552 (1935).

⁶¹H. Pickett and H. Strauss, *J. Am. Chem. Soc.* **92**, 7281 (1970).

⁶²K. Kudin and A. Dymarsky, *J. Chem. Phys.* **122**, 224105 (2005).

⁶³G. Kneller, *J. Chem. Phys.* **128**, 194101 (2008).

⁶⁴L. Maragliano, E. Vanden-Eijnden, and B. Roux, *J. Chem. Theory Comput.* **5**, 2589 (2009).

⁶⁵J. L. Adelman and M. Grabe, *J. Chem. Phys.* **138**, 044105 (2013).

⁶⁶L. Maragliano, G. Cottone, G. Ciccotti, and E. Vanden-Eijnden, *J. Am. Chem. Soc.* **132**, 1010 (2010).

⁶⁷V. Ovchinnikov, M. Karplus, and E. Vanden-Eijnden, *J. Chem. Phys.* **134**, 085103 (2011).

⁶⁸W. Kabsch, *Acta Cryst.* **A34**, 827 (1978).

⁶⁹A. Sayvetz, *J. Chem. Phys.* **7**, 383 (1939).

⁷⁰S. Ferigle and A. Weber, *Am. J. Phys.* **21**, 102 (1953).

⁷¹E. Wilson, Jr., J. Decius, and P. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).

- ⁷²D. R. Herschbach, H. S. Johnston, and D. Rapp, *J. Chem. Phys.* **31**, 1652 (1959).
- ⁷³S. Boresch and M. Karplus, *J. Chem. Phys.* **105**, 5145 (1996).
- ⁷⁴See supplementary material at <http://dx.doi.org/10.1063/1.4871685> for the alanine dipeptide test case, for discussions of the electrostatic and solvation energies and of rotational contributions to the free energy of the β -hairpin, and for commitment analysis of the ZTS path.
- ⁷⁵E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **130**, 194101 (2009).
- ⁷⁶A. Brünger, C. Brooks, and M. Karplus, *Chem. Phys. Lett.* **105**, 495 (1984).
- ⁷⁷B. Brooks, C. Brooks III, A. Mackerell, Jr., L. Nilsson, R. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch *et al.*, *J. Comput. Chem.* **30**, 1545 (2009).
- ⁷⁸J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, and K. Schulten, *J. Comput. Chem.* **26**, 1781 (2005).
- ⁷⁹H. Fukunishi, O. Watanabe, and S. Takada, *J. Chem. Phys.* **116**, 9058 (2002).
- ⁸⁰G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, *J. Am. Chem. Soc.* **128**, 13435 (2006).
- ⁸¹M. Cecchini, S. Krivov, M. Spichty, and M. Karplus, *J. Phys. Chem. B* **113**, 9728 (2009).
- ⁸²M. Spichty, M. Cecchini, and M. Karplus, *J. Phys. Chem. Lett.* **1**, 1922 (2010).
- ⁸³V. Ovchinnikov, M. Cecchini, and M. Karplus, *J. Phys. Chem. B* **117**, 750 (2013).
- ⁸⁴B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- ⁸⁵A. MacKerell, Jr., M. Feig, and C. Brooks III, *J. Comput. Chem.* **25**, 1400 (2004).
- ⁸⁶U. Haberthür and A. Caflisch, *J. Comput. Chem.* **29**, 701 (2008).
- ⁸⁷G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.* **7**, 95 (1963).
- ⁸⁸G. J. Kleywegt and T. A. Jones, *Structure (London)* **4**, 1395 (1996).
- ⁸⁹R. W. Hooft, C. Sander, and G. Vriend, *Comput. Appl. Biosci.* **13**, 425 (1997).
- ⁹⁰R. Best, X. Zhu, J. Shim, P. Lopes, J. Mittal, M. Feig, and A. MacKerell, Jr., *J. Chem. Theor. Comput.* **8**, 3257 (2012).
- ⁹¹R. Zhou, *Proteins: Struct. Funct. Genet.* **53**, 148 (2003).
- ⁹²F. Blanco, G. Rivas, and L. Serrano, *Nat. Struct. Biol.* **1**, 584 (1994).
- ⁹³E. Schreiner, L. Trabuco, P. Freddolino, and K. Schulten, *BMC Bioinformatics* **12**, 190 (2011).
- ⁹⁴J.-P. Ryckaert, G. Ciccotti, and H. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- ⁹⁵H. De Loof, L. Nilsson, and R. Rigler, *J. Am. Chem. Soc.* **114**, 4028 (1992).
- ⁹⁶B. Dickson, H. Huang, and C. B. Post, *J. Phys. Chem. B* **116**, 11046 (2012).
- ⁹⁷M. Searle and D. Williams, *J. Am. Chem. Soc.* **114**, 10690 (1992).
- ⁹⁸P. Gilli, V. Ferretti, G. Gilli, and P. Borea, *J. Phys. Chem.* **98**, 1515 (1994).
- ⁹⁹K. Frederick, M. Marlow, K. Valentine, and J. Wand, *Nature (London)* **448**, 325 (2007).
- ¹⁰⁰J. Dunitz, *Science* **264**, 670 (1994).
- ¹⁰¹J. Dunitz, *Chem. Biol.* **2**, 709 (1995).
- ¹⁰²H. Yu, L. Ma, Y. Yang, and Q. Cui, *PLoS Comput. Biol.* **3**, 0214 (2007).
- ¹⁰³H. Huang, E. Ozkirimli, and C. Post, *J. Chem. Theory Comput.* **5**, 1304 (2009).
- ¹⁰⁴A. A. Golosov, J. J. Warren, L. S. Beese, and M. Karplus, *Structure* **18**, 83 (2010).
- ¹⁰⁵D. 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).
- ¹⁰⁶K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, *Science* **334**, 517 (2011).
- ¹⁰⁷T. J. Lane, G. R. Bowman, K. Beauchamp, V. A. Voelz, and V. S. Pande, *J. Am. Chem. Soc.* **133**, 18413 (2011).
- ¹⁰⁸F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19011 (2009).
- ¹⁰⁹R. Radhakrishnan and T. Schlick, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5970 (2004).
- ¹¹⁰D. de Jong, G. Singh, W. Bennett, C. Arnarez, T. Wassenaar, L. Schäfer, X. Periole, D. Tieleman, and S. Marrink, *J. Chem. Theory Comput.* **9**, 687 (2013).
- ¹¹¹T. Veitshans, D. Klimov, and D. Thirumalai, *Fold. Des.* **2**, 1 (1997).
- ¹¹²M. Levitt and A. Warshel, *Nature* **253**, 694 (1975).
- ¹¹³B. Brooks, D. Janežič, and M. Karplus, *J. Comput. Chem.* **16**, 1522 (1995).