Minimal Folding Pathways for Coarse-Grained Biopolymer Fragments

Ali R. Mohazab and Steven S. Plotkin
Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada

ABSTRACT The minimal folding pathway or trajectory for a biopolymer can be defined as the transformation that minimizes the total distance traveled between a folded and an unfolded structure. This involves generalizing the usual Euclidean distance from points to one-dimensional objects such as a polymer. We apply this distance here to find minimal folding pathways for several candidate protein fragments, including the helix, the β-hairpin, and a nonplanar structure where chain noncrossing is important. Comparing the distances traveled with root mean-squared distance and mean root-squared distance, we show that chain noncrossing can have large effects on the kinetic proximity of apparently similar conformations. Structures that are aligned to the β-hairpin by minimizing mean root-squared distance, a quantity that closely approximates the true distance for long chains, show globally different orientation than structures aligned by minimizing root mean-squared distance.

INTRODUCTION

The mechanism by which a biopolymer folds has been a subject of long-standing interest. Some of the questions often focus on identifying structural features of the transition state or nucleus (1–11), which is in many formulations characterized through the commitment or splitting probability, probably first introduced by Onsager in the context of ion-pair recombination (12). Another related question of much interest is how to characterize coordinates that best represent progress in the reaction (13–27). Questions often circulate around what parameter(s) or principle component-like motions might best correlate with splitting probability or probability of folding before unfolding. Finding the folding trajectories that are the most energetically downhill is also of interest (28,29).

Useful order parameters have a simple geometric interpretation. This has led, for example, to the common use of the fraction of native contacts Q (3,20,22,30–33), which can be locally or globally defined; root mean-square distance or deviation (RMSD) between structures (34–37); structural overlap parameter χ (38–40); Debye-Waller factors (41,42); or fraction of correct dihedral angles (33).

While the utility of simple order parameters is indisputable, it is easy to see that even for simple structures they can lead to inaccurate measurements of native proximity. For example, a β-hairpin that is only slightly expanded beyond the range of its hydrogen bonds is essentially committed to fold, but would have a Q value near zero. Comparing two conformations of a piece of polymer chain that crosses either over itself or under itself would give a χ-parameter or RMSD that could be quite small. The amount of motion the polymer would have to undergo to transform from one conformation to the other, however, respecting the noncrossing constraint, would have to be comparably large.

Here we propose a new candidate for an order parameter to capture the complexities of biomolecule folding. Specifically, we consider a generalization of the conventional notion of distance to polymeric objects. This distance depends only on the geometry of the initial and final configurations. In two previous articles (43,44) we have introduced the formalism for generalizing the standard variational definition of distance to higher dimensional objects such as polymers or membranes. Here we apply this formalism to fragments of coarse-grained protein backbone structures.

To obtain a numerical solution for practical applications, the polymer chain must be discretized. Then we are interested in the cumulative distance that all beads on the chain must travel to convert chain conformation A characterized by $\{\mathbf{r}_{\mathrm{A}i}\}$ with $i = 1, 2. . .N$, to conformation B characterized by $\{\mathbf{r}_{\mathrm{B}i}\}$. We model the chains as having $N-1$ links of length Δs which are incompressible and inextensible, so that $(\mathbf{r}_{i+1} - \mathbf{r}_i)^2 = \Delta s^2 \equiv 1$. This requires that we introduce a Lagrange multiplier $\lambda_{i,i+1}$ for each link i, $1 \le i \le N-1$.

The transformation from A to B may be written in terms of an artificial time parameter t as $\{\mathbf{r}_i(t)\}$, with boundary conditions, or in keeping with the language of time, in both initial and final conditions:

$$\{\mathbf{r}_i(0)\} = \{\mathbf{r}_{\mathrm{A}i}\} \text{ and } \{\mathbf{r}_i(T)\} = \{\mathbf{r}_{\mathrm{B}i}\}. \quad (1)$$

The distance $\mathcal{D}$ traveled in such a transformation can be written as a functional $\mathcal{D}[\{\mathbf{r}_i(t)\}, \{\dot{\mathbf{r}}_i(t)\}]$, or $\mathcal{D}[\mathbf{r}_i, \dot{\mathbf{r}}_i]$ for short (43,44)

$$\mathcal{D}[\mathbf{r}_i, \dot{\mathbf{r}}_i] = \int_0^T dt\, L(\mathbf{r}_i, \dot{\mathbf{r}}_i), \quad (2)$$

where the integrand L can be thought of as an effective Lagrangian:

$$\mathcal{L}(\mathbf{r}_i, \dot{\mathbf{r}}_i) = \sum_{i=1}^{N} \left(\sqrt{\dot{\mathbf{r}}_i^2} - \frac{\lambda_{i,i+1}}{2}((\mathbf{r}_{i+1} - \mathbf{r}_i)^2 - 1) \right). \quad (3)$$

Equations 2 and 3 are simply a way to write the distance traveled by all the beads as a variational problem. We can

Submitted April 9, 2008, and accepted for publication June 10, 2008.

Address reprint requests to Steven S. Plotkin, Tel.: 604-822-8813; E-mail: steve@physics.ubc.ca.

Editor: Angel E. Garcia.

0006-3495/08/12/5496/12 $2.00

doi: 10.1529/biophysj.108.135046

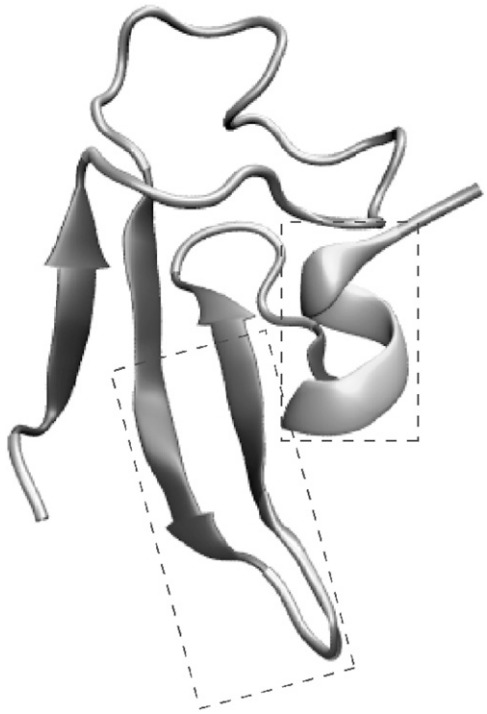

FIGURE 1 Residues 99–153 in regulatory chain B of Aspartate Carbamoyltransferase (47) (PDB code 1AT1) are chosen for analysis. From this domain we select three fragments for investigation. Two are outlined in dashed boxes: β-hairpin residues 126–137, and α-helix residues 147–151. The strand 1-turn-strand-2 tertiary motif, residues 101–130, is also used investigate the importance of noncrossing.

then minimize this functional to find the variational minimum distance transformation between two biopolymer fragments.

Numerically, the distance $\mathcal{D}$ in Eq. 2 is in units of link-length-squared, i.e., it is one-dimension-of-length higher than the boundary conditions, just as the distance between points has dimension 1.

The minimal distance transformation between an initial polymer conformation A and the folded or native conformation N can be thought of as an optimal folding pathway that is the most direct route from A to N. Of course, the actual trajectory is a stochastic one. It is interesting to ask whether the typical or average dynamical trajectory resembles the minimal one after suitable averaging, but we do not answer this question here. Interaction energies in the system will certainly modify the weights of reactive trajectories, making some trajectories preferred over others. On the other hand, much of the folding mechanism is thought to be insensitive to specific sequence details (45), and depends more on the geometry of the native structure and its resultant topology of interactions (46).

A direct application of minimal folding path to a full protein is an important future goal. In this article, we take a more bottom-up, modular approach, and apply the minimal distance transformation to various representative protein fragments. In particular, we investigate the minimal folding pathways for a β-sheet, an α-helix, and an overpass-underpass problem, where chain noncrossing is important.

METHODS

Optimal pathways

We refer to the transformation between structures A and N that minimizes the distance functional in Eq. 2 as the minimal transformation or optimal folding pathway. This transformation is found by extremizing $\mathcal{D}$:

$$\delta\mathcal{D}[\mathbf{r}_i(t), \dot{\mathbf{r}}_i(t)] = 0. \tag{4}$$

The solutions to Eq. 4 are a set of differential equations for $\{\mathbf{r}_i(t)\}$. Defining $\dot{x} \equiv dx/dt$, $\mathbf{r}_{i/j} \equiv \mathbf{r}_i - \mathbf{r}_j$, and $\hat{\mathbf{v}}_i$ as the unit velocity vector $\mathbf{v}_i/|\mathbf{v}_i|$, the solutions to Eq. 4 are written as

$$\dot{\hat{\mathbf{v}}}_1 + \lambda_{12}\,\mathbf{r}_{2/1} = 0, \tag{5a}$$

$$\dot{\hat{\mathbf{v}}}_2 - \lambda_{12}\,\mathbf{r}_{2/1} + \lambda_{23}\,\mathbf{r}_{3/2} = 0, \tag{5b}$$

$$\vdots$$

$$\dot{\hat{\mathbf{v}}}_N - \lambda_{N-1,N}\,\mathbf{r}_{N/(N-1)} = 0. \tag{5c}$$

In practice, piecewise extremal solutions are observed consisting of sequential rotations and translations (44). This leads to so-called corner conditions at the interface between two pieces of the extremal solution. At the interface, the corner conditions demand that

$$\lim_{\epsilon\to 0} \hat{\mathbf{v}}_i(t-\epsilon) = \lim_{\epsilon\to 0} \hat{\mathbf{v}}_i(t+\epsilon) \tag{6}$$

at the interface between piecewise solutions to Eqs. 5a–5c, for the trajectory of bead i to be extremal. This simply means that the trajectory of bead i cannot suddenly change direction. If it did, the resulting trajectory with cusp could always be shortened by truncating the cusp, or cutting the corner. One possible exception to this rule is if at some part of the extremal trajectory the velocity of bead i is zero (the point is at rest). Then its direction $\hat{\mathbf{v}}$ is undefined, and can, in principle, change at a later time. Another possible exceptional case is when there are external forces on the link, which can be induced for example by noncrossing constraints, discussed below.

From Eq. 5a, there are three solutions for the end bead of the chain $\mathbf{r}_1$, with analogous solutions for $\mathbf{r}_N$ in Eq. 5c.

1. If $\lambda_{12} = 0$, $\dot{\hat{\mathbf{v}}}_1 = 0$, and straight-line motion of the end point results.
2. If $\lambda \neq 0$, the velocity of the end point is orthogonal to the link, which we can see by taking the dot product of Eq. 5a with $\mathbf{v}_1$, i.e., $\mathbf{v}_1 \cdot \mathbf{r}_{2/1} = 0$. The result is pure rotational motion of the bead.
3. The end point can remain at rest while other parts of the chain move. This can be seen by writing out the time-derivative in Eq. 5a,

$$\mathbf{v}_1^2\dot{\mathbf{v}}_1 - (\mathbf{v}_1 \cdot \dot{\mathbf{v}}_1)\mathbf{v}_1 = -\lambda_{12}|\mathbf{v}_1|^3\mathbf{r}_{2/1},$$

which has the trivial solution $\mathbf{v}_1 = 0$.

By piecing together rotations and subsequent translations, we can find extremal solutions for the transformation from structure A to N. It was shown in Mohazab and Plotkin (44) that these solutions are also minimal. The intermediate conformations have solitonic kinks that propagate along the chain.

Representative protein fragments

For a protein domain to which we apply our methods, we choose residues 99–153 in regulatory chain B of Aspartate Carbamoyltransferase (47) (PDB code 1AT1, see Fig. 1). From this domain, we select three fragments for investigation, as representatives of some commonly found secondary and tertiary structures:

1. The β-hairpin containing β-strands 2 and 3, residues 126–137.
2. The C-terminal α-helix, residues 147–151.
3. The β-strand 1-turn-strand 2 tertiary motif, residues 101–130.

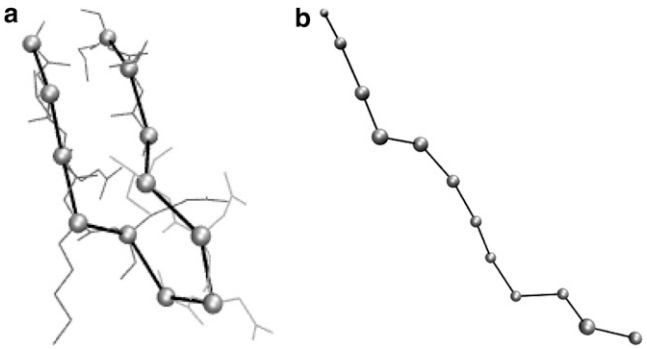

FIGURE 2 (*a*) β-hairpin fragment, with all-atom and coarse-grained C_α representations superposed. (*b*) The extended initial state.

We investigate an overpass/underpass problem for a simplified version of segment 3 for which chain noncrossing is important.

The polymer fragments are coarse-grained by taking the C_α atom to represent each residue. The C_α–C_α distances in our fragments are sharply peaked: $|\mathbf{r}_{i+1/i}| = (3.81 \pm 0.04)$ Å. We do not change the numbers present in the PDB structure: they are held fixed during the transformation.

We investigate the minimal distance transformations between extended states of polymer and the above secondary structures. Extended states are constructed as follows. For the β-hairpin, we rotate the chain about the positions of $C_\alpha(132)$ and $C_\alpha(133)$ so that the initial state is an extended linear strand (Fig. 2 *b*). For the α-helix, we take the simplified case of a straight line for the initial condition.

For the over/under problem we imagine a scenario where the β-sheet in Fig. 3 *a* is unformed, and the polymer chain involved in the turn has crossed under rather than over β-stand 2. The two configurations have the opposite sense, in that the chain must cross over itself (or go over the top or the bottom of the structure) to form the correct tertiary structure (Fig. 3 *b*). Alternatively, β-strands 2 and 3 in Fig. 1 may cross over β-strand 1 to solve the underpass-overpass problem, but this would involve larger-scale motion, that is, a larger distance traveled.

A stereo view of initial and final states for such a scenario is shown in Fig. 3 *b*. We ask: What is the minimal distance pathway for conversion between these two structures? To make the problem more amenable to analysis, we simplify the structures in the spirit of lattice models, as shown in Fig. 3 *c*. The initial and final conditions are regular and symmetric, but intermediate configurations can be anywhere so long as they are consistent with the constraints of constant link length and noncrossing (i.e., they can be off-lattice).

Construction of minimal pathways

Minimal folding trajectories are constructed by the recipe described in Mohazab and Plotkin (44) (Fig. 4). The basic recipe is as follows. First we take the coordinate of one C_α residue, say $\mathbf{r}(C_{\alpha i})$ in the unfolded conformation, then we imagine rotating $\mathbf{r}(C_{\alpha i})$ about $\mathbf{r}(C_{\alpha(i-1)})$. The protein backbone is treated approximately as a freely jointed chain to carry out this procedure. All possible rotations of $C_{\alpha i}$ about $C_{\alpha(i-1)}$ form a sphere of radius $|\mathbf{r}(C_{\alpha i}) - \mathbf{r}(C_{\alpha(i-1)})|$. A cone is drawn from the final position of $C_{\alpha i}$, i.e., $\mathbf{r}^{\mathrm{FOLDED}}(C_{\alpha i})$ in the folded structure, to be tangent to this sphere. In general, one particular direction will have the minimal amount of rotation before proceeding in a straight line to $\mathbf{r}^{\mathrm{FOLDED}}(C_{\alpha i})$. The arc of the great circle along this direction is then chosen as part of the minimal trajectory for residue *i*.

RMSD and MRSD

In the limit of long polymer chains and in the absence of noncrossing, the distance accumulated by rotation of each link before translating gives a negligible contribution to the total distance, and the total distance traveled converges to the chain length L times the mean root-square distance (MRSD), i.e., for two structures A and B,

$$\lim_{N\to\infty} \mathcal{D} = L \times \frac{1}{N}\sum_i \sqrt{(\mathbf{r}_{Bi} - \mathbf{r}_{Ai})^2} = L \times (\mathrm{MRSD}). \quad (7)$$

It can be shown that the MRSD is always less than the RMSD defined by $\sqrt{N^{-1}\sum_i (\mathbf{r}_{Ai} - \mathbf{r}_{Bi})^2}$, and often used for structural comparison. Which of these quantities provides more accuracy for structural alignment is still an open question, although the MRSD may be less sensitive to large fluctuations of a subset of points. To investigate the sensitivity of MRSD versus RMSD, note that the change in RMSD with respect to moving one residue an amount $\delta\mathbf{r}_{Ai}$ is

$$\left|\frac{\delta(RMSD)}{\delta\mathbf{r}_{Ai}}\right| \approx \frac{1}{N}\frac{|\mathbf{r}_{Ai} - \mathbf{r}_{Bi}|}{RMSD},$$

while the change in MRSD with respect to moving one residue an amount $\delta\mathbf{r}_{Ai}$ is

$$\left|\frac{\delta(MRSD)}{\delta\mathbf{r}_{Ai}}\right| \approx \frac{1}{N}.$$

So if residue i has a structural discrepancy larger than the average as measured by RMSD, changes in RMSD with respect to this residue's position will be larger than those for MRSD.

Unfolded conformations were aligned to folded structures by minimizing MRSD and RMSD, and minimal transformations constructed for these conformation pairs. For the β-hairpin, the conformation pairs were observed to be globally different depending on whether the alignment cost function was MRSD or RMSD.

RESULTS

β-hairpin

We coarse-grain the fragment containing residues 126–137 by considering only the C_α atoms (see Fig. 2 *a*). We consider folding to this structure from an extended state. The extended state is obtained by two rotations about residues 132 and 133, which extend the hairpin out to a quasilinear strand (the extended state in Fig. 2 *b*). This initial extended state is aligned to the final structure in four different ways:

1. One strand of the hairpin is directly aligned to the corresponding residues of the extended state (Fig. 5, *a* and *b*),
2. The center links of the hairpin and extended state are directly aligned to each other (Fig. 5 *c*),
3. The initial position/orientation of the extended state is found by minimizing the MRSD between the two coarse-grained C_α structures (hairpin and extended state) in Fig. 2, *a* and *b* (Fig. 5 *d*, *blue extended strand*), and
4. The initial position/orientation of the extended state is found by minimizing the RMSD between the two coarse-grained C_α structures (hairpin and extended state) in Fig. 2, *a* and *b* (Fig. 5 *d*, *teal extended strand*).

From these initial states, we have found minimal folding trajectories consisting of rotations and subsequent translations of the residues (or vice versa) as described in Methods.

To gain intuition for the transformations from the MRSD and RMSD aligned structures, we also considered minimal trans-

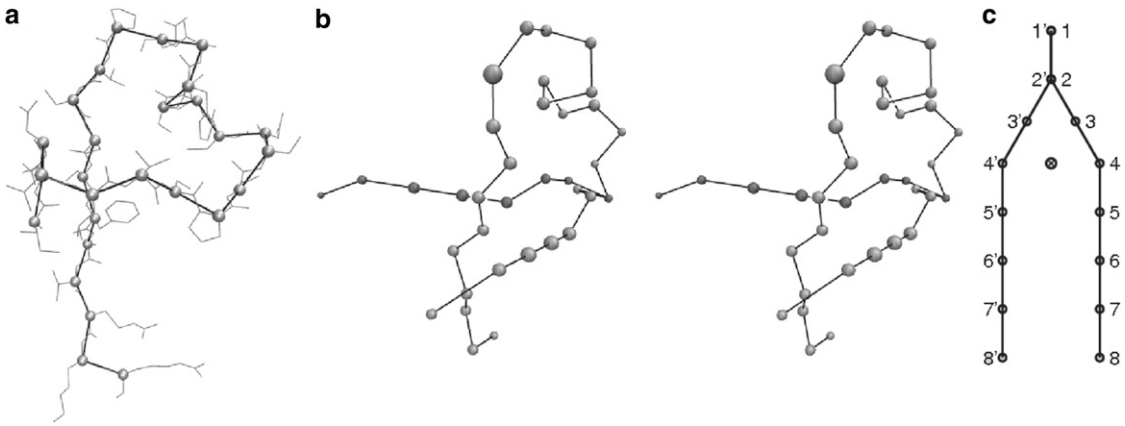

FIGURE 3 (*a*) Residues 101–130 of Aspartate Carbamoyltransferase can be taken as an example of an overpass/underpass problem where chain noncrossing is important. (*b*) Conformation of the segment in panel *a* with the β-sheet unformed. Both initial and final structures (with opposite over/under sense) are superposed in this stereo view. (*c*) A simplified model to capture the essence of the underpass-overpass problem. Both initial and final states are shown as viewed from above. Residues 1–8 must transform to residues $1'-8'$, but cannot pass through the obstacle marked with a circled *X*, representing a long piece of polymer normal to the plane of the figure.

formations from an idealized straight-line structure to an idealized β-hairpin, whose initial and final states are shown in Fig. 5 *e*.

The distances for all the β-hairpin transformations, along with numbers for the RMSD and MRSD for the same transformations, are given in Table 1.

The resulting transformations for the above boundary conditions are shown in Fig. 5, *a–c*, and *f–i*. As described in Methods, the minimal folding pathways proceed by forming kinks or solitonic-like waves that propagate along the backbone. The soliton-like object consists of a rotation of a bead until the link containing that bead reaches a critical angle. The bead subsequently translates until it reaches its final position.

For the idealized straight-line to β-hairpin transformation, the MRSD and RMSD aligned structures are globally different (Fig. 5 *e*). The MRSD between the two aligned straight-line structures is 15.39 Å, larger than the MRSD of either structure to the folded hairpin state (Table 1). The transformation from the RMSD-aligned line involves predominantly straight-line motion from the line to the hairpin (Fig. 5 *f*). Only ~0.1% of the distance corresponds to rotational motion. The transformation from the MRSD-aligned line involves both rotations and translations, as shown in Fig. 5 *g*. This gives the MRSD-aligned pair a distance only marginally smaller (0.4%) than the RMSD-aligned pair (Table 1), even though the transformations have different initial states and very different character.

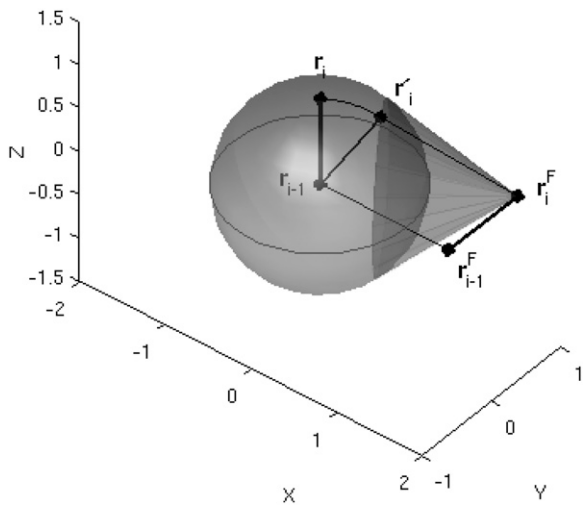

FIGURE 4 Illustration of the general recipe for obtaining minimal pathways (see Methods).

For the real β-hairpin and extended state, the transformations are reminiscent of the ideal case. The MRSD and RMSD aligned structures are globally different, as shown in Fig. 5 *d*. The MRSD between the two aligned extended structures is 9.83 Å, which is again larger than the MRSD of either structure to the folded hairpin state (Table 1). The MRSD-aligned pair has a distance 17% different than the ideal case and the RMSD-aligned pair has a distance 23% different than the ideal case. Fig. 5, *h* and *i*, depict the transformations for RMSD- and MRSD-aligned pairs, respectively. For the real β-hairpin, the RMSD-aligned extended state has a smaller distance than the MRSD-aligned extended state by ~5%, i.e., the scenario present in the idealized case is reversed, somewhat surprisingly. This indicates that the aligned structures obtained by minimizing the actual distance need not resemble those structures obtained by either the RMSD or MRSD alignments. An alignment algorithm for general structures using distance $\mathcal{D}$ as a cost function is a nontrivial problem that we reserve for future work.

We note that the above transformations will not all have the same energy gain as they fold. Transformations in Fig. 5, *a* and *c*, are similar in the main to the energetically driven zippering and assembly mechanisms of conformational search proposed by Ozkan et al. (48). A folding pathway similar to the transformation in Fig. 5 *b* would not have con-

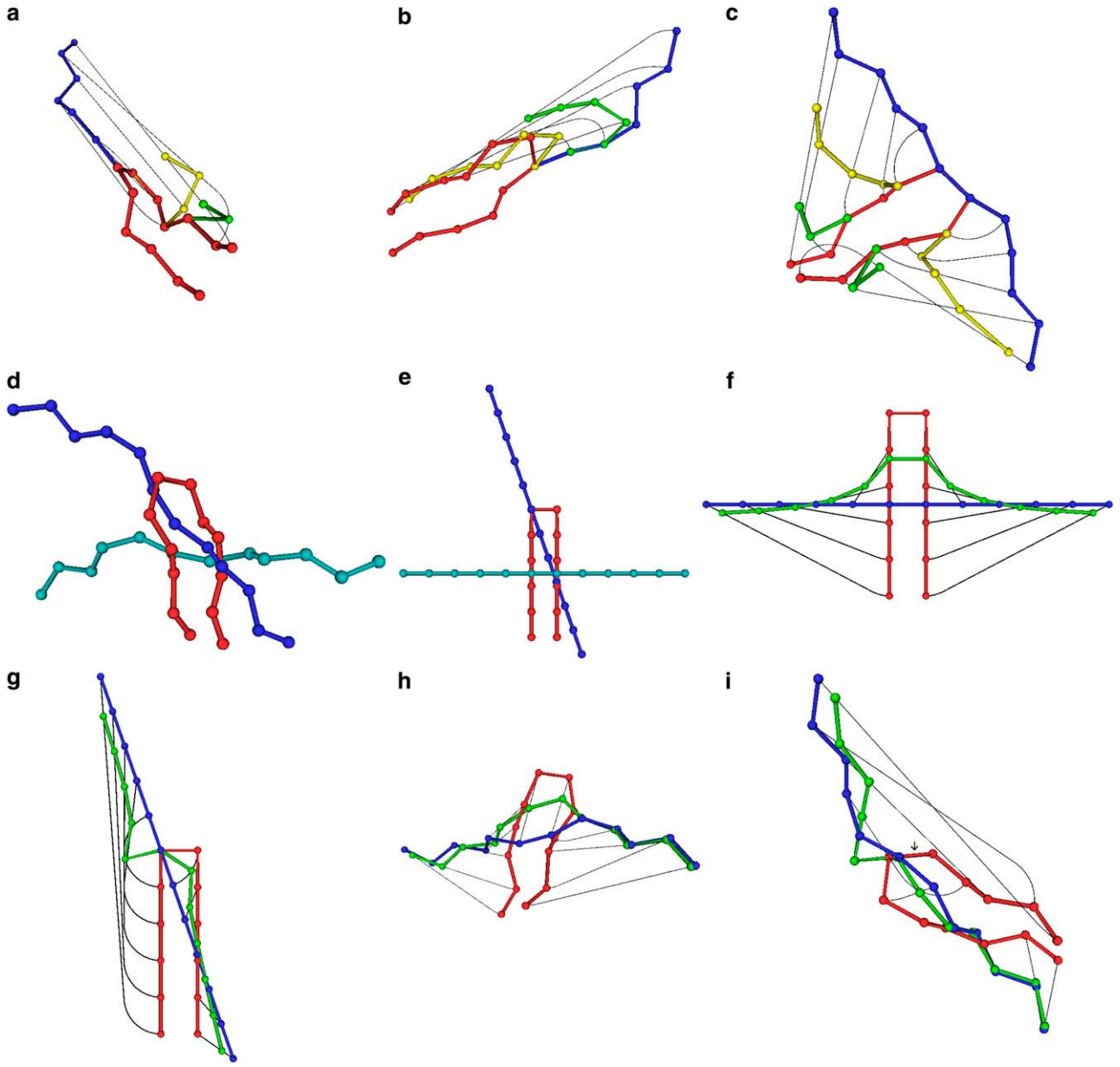

FIGURE 5 Minimal transformations to the β-hairpin. Distances are given in Table 1. (*a*) Folding pathway in which one strand of the hairpin can be thought of as peeling away by rotations of the links to various critical angles, which are then followed by subsequent translations into their final positions. (*b*) A minimal pathway that can be thought of as involving kink propagation or peeling away from the extended strand, followed by translation of the links into their final positions in the β-hairpin. (*c*) A zippering mechanism, in which we have aligned the middle link of the hairpin and sought the minimal distance transformation. The distance here is somewhat larger than the distance for the transformations in panels *a* and *b*. (*d*) The extended strand is aligned to the β-hairpin by minimizing RMSD (*blue*), or minimizing MRSD (*teal*). (*e*) Idealized version of the extended strand and β-hairpin. The extended strand is again aligned to the β-hairpin by minimizing RMSD (*blue*), or minimizing MRSD (*teal*). (*f*) Transformation for the idealized β-hairpin, for RMSD-aligned structures. Initial state is blue, final state is red, and intermediate state is in green. (*g*) Transformation for the idealized β-hairpin, for MRSD-aligned structures. (*h*) RMSD-aligned transformation between the extended strand (*blue*) and β-hairpin (*red*). An intermediate state is shown in green. (*i*) MRSD-aligned transformation between the extended strand (*blue*) and β-hairpin (*red*). An intermediate state is shown in green. The small arrow points to a link with an somewhat unconventional transformation, which is discussed in the Appendix.

current energy gain and so would be less likely thermodynamically. To implement the transformation shown in Fig. 5 *c*, the construction described in Methods above and shown in the figure is only approximately correct, to ~1%. To find an exact minimal solution involves generalizing the methodology to allow for concurrent rotations of two links about a central axis, as described in more detail in Sections 4 and 5 of Mohazab and Plotkin (44).

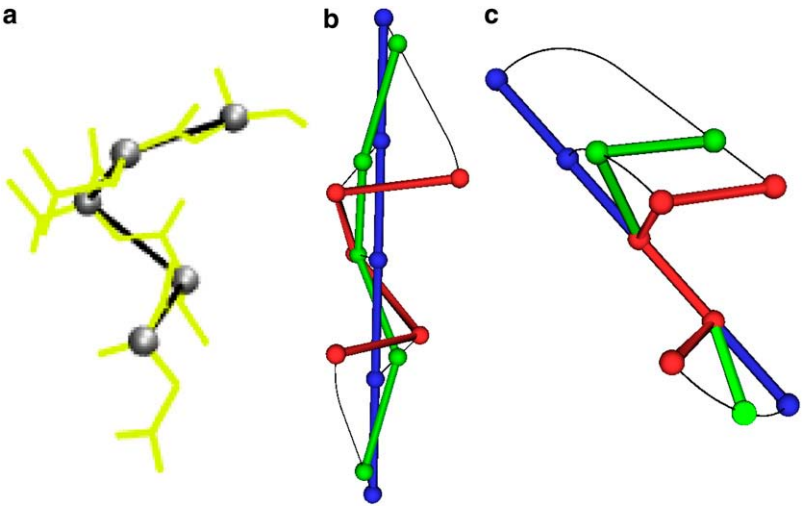

FIGURE 6 (*a*) Single α-helix of five residues 147–151 taken from PDB 1AT1. (*b*) Minimal pathway to fold the α-helix (*red*), from a straight line initial state which has been aligned by minimizing MRSD (shown in *blue*, see text for description). A conformation partway though the transition is shown in green. (*c*) Minimal pathway to fold the helix from a straight-line initial conformation with its second link directly aligned to the second link of the helix. Distances for both transformations are given in Table 1.

α-helix

We coarse-grain the helical fragment containing residues 147–151 by considering only the $C\alpha$ atoms (see Fig. 6 *a*).

We consider folding to this structure from an extended state. The extended state is taken for simplicity to be a straight line. Of course more realistic extended conformations could be taken, but would give minor quantitative corrections to the numbers we obtain.

We consider two different initial conditions for the straight line, one where link 2 is exactly aligned with link 2 of the α-helix (Fig. 6 *c*), and one where the straight line is aligned to the helix by minimizing the MRSD. This initial condition is such that the straight line threads the helix (Fig. 6 *b*). The aligned unfolded structure obtained by minimizing RMSD is similar in this case: the MRSD between the two aligned structures is only 1.53 Å.

From these initial states, we found minimal folding trajectories consisting of rotations and subsequent translations from the straight-line conformation to the helix.

Fig. 6 *b* shows a minimal folding pathway to the α-helix. An intermediate conformation (partway through the transition) is shown in green.

The distance traveled after minimizing MRSD is indeed less than the distance after alignment of one link. For both of these transformations, the distances traveled per residue are less than the corresponding distance per residue for the β-hairpin transformations.

Crossover structure

The fact that the polymer chain cannot cross itself is represented by inequality constraints in the equations of motion. We introduce the methods for solution of variational problems with inequality constraints in the Appendix. The upshot is that the minimal distance problem is a free problem until a residue on the chain touches the obstacle. At that point the residue is constrained to be on the surface of the obstacle and the trajectory is defined accordingly. Eventually the particle or residue leaves the surface, and the problem becomes a free problem once again, as the particle moves to its final position. The transformation is then piecewise, consisting of three pieces, and at the interface between the pieces, the corner conditions (Eq. 6) must hold.

The initial and final conditions of an idealized noncrossing chain are shown in Fig. 3 *c*. In our problem of chain noncrossing, the obstacle is an effectively infinite line, normal to the plane of Fig. 3 *c* (marked by a circled *X*), so residues only need to touch that point before proceeding to their final position. In this treatment residues are treated asymmetrically, in that one part of the chain has steric hindrance along bonds, while another only has steric hindrance for the masses or beads at the termini of bonds. This approximation is assumed to simplify the transition, and because the resulting distance only differs by a small finite size-effect from the distance obtained by employing links for all parts of the chain.

We found a solution that fully satisfies the Euler-Lagrange (EL) equations Eqs. 5a–5c, and corner conditions satisfy Eq. 6. According to our previous work (44), this class of

TABLE 1 Values of the distance for various protein backbone fragments, as compared to other metrics

Backbone conformation	Figure	$\mathcal{D}/(N\ell)$*	RMSD	MRSD
β-Hairpin (half-aligned)	5 *a*	10.372	15.538	9.926
β-Hairpin (half-aligned)	5 *b*	10.372	15.538	9.926
β-Hairpin (zipper)	5 *c*	12.787	13.560	11.317
β-Hairpin (RMSD-aligned)	5 *h*	9.749	10.501	9.730
β-Hairpin (MRSD-aligned)	5 *i*	10.277	12.681	9.412
Ideal β-hairpin (RMSD-aligned)	5 $f^{\dagger}$	12.25	13.24	12.24
Ideal β-hairpin (MRSD-aligned)	5 $g^{\dagger}$	12.18	16.31	11.27
α-Helix (MRSD aligned)	6 *b*	3.595	3.954	3.577
α-Helix (1-link aligned)	6 *c*	4.675	5.805	4.233
Over/under (noncrossing)	$7^{\dagger}$	13.991	6.173	5.239

*Distance $\mathcal{D}$ is divided by N times the link length ℓ, so that all quantities in the table have units of Å.

$^{\dagger}\mathcal{D}$ is put in the same units as the above transformations, i.e., we take $\ell = 3.81$ for the link length.

solution is at least a local minimum. It involves the propagation of a kink starting at the end of the chain, in which the chain proceeds snakelike over the obstacle and then back down to its final position, and so is intuitively reasonable. The distance is given in Table 1, along with the RMSD and MRSD. In cases where noncrossing is important, the distance $\mathcal{D}$ will be significantly greater than either RMSD or MRSD.

The transformation starts by a rotation of link 7-8 about the point 7, until a critical angle $\pi/2$ is reached. Residue 8 subsequently translates to the crossover point O. Immediately as it starts translating, link 6-7 rotates about point 6 (Fig. 7 *a*) and residue 7 rotates to its critical angle of $\pi/2$. The process repeats until link 5-6 rotates to an angle of $\pi/6$, at which point residue 8 touches the obstacle (Fig. 7 *b*).

At this point, residue 8, which is touching a nondifferentiable (nonsmooth) surface, may violate corner conditions for the reasons discussed in the Appendix. Residue 8 moves horizontally to the left while residue 7 moves vertically, so the end points of the link slide in orthogonal directions (Fig. 7 *c*). After this part of the transformation is complete, the chain is in the configuration shown in Fig. 7 *d*.

At this point, link 4-5 begins to rotate, and this sets up a cascade of motions throughout the chain. Residue 8 slides vertically downward, residue 7 slides horizontally to the left, and residue 6 slides vertically upward (Fig. 7 *e*). Note that residue 8 appears to violate corner conditions in the opposite sense of residue 7. These violations are again due to the influence of the crossover constraint.

When link 4-5 has rotated to $\pi/6$, link 6-7 is horizontal and link 7-8 is vertical (Fig. 7 *f*). As 4-5 continues to rotate, residues 7 and 8 proceed vertically downward in Fig. 7 *g*, while residue 6 moves left horizontally, until the conformation in Fig. 7 *h* is reached when link 4-5 has finished its rotation to $\pi/2$.

At this point link 3-4 begins to rotate about position 3, moving residue 4 to the noncrossing position O, while the rest of the chain shifts downward vertically in the Fig. 7, *i* and *j*.

Finally residue 3 rotates about position 2 while residue 4 translates in a straight line to its final position, and all other residues translate downwards (Fig. 7, *k* and *l*).

This completes the transformation. Note again that the distance in Table 1 is much larger than either the RMSD or MRSD. A second transformation is obtained by time-reversing the above solution, and swapping the right and left branches of the structure that serve as initial and final conditions.

DISCUSSION AND CONCLUSIONS

In this article, we have applied the general theory of distance between high dimensional objects to find the minimal folding pathways for protein fragments. We consider this to be a first step in building up ever-larger fragments to eventually look at the distance as an order parameter for the folding of a whole biomolecule.

We investigated the minimal folding pathway for a helix, a β-hairpin, and a structure involving a crossover where the integrity of the chain is essential in determining the minimal transformation.

The noncrossing problem has the largest distance per residue of all conformations considered. Not surprisingly, the α-helix has the shortest. It is an interesting future question to address the consequence of the distance from a random structure to a folded structure on its folding rate.

We have made several approximations in our model. In our analysis of minimal distance trajectories, we have not accounted for the steric excluded volume of the side chain and backbone degrees of freedom that have been coarse-grained out. It is possible to account for this in principle by applying the methods described in the Appendix. We take the trajectories derived here as a first approximation to the more fully constrained problem.

Another modification that must be considered is the range of allowed angles between consecutive triples of C_α residues. While sharp kinks in our transformations were the exception rather than the rule, we have assumed in our analysis that the full range of angles is allowable. The coarse-graining procedure does give greater flexibility for the resulting chain because there are six backbone bonds per C_α triple; however, a more thorough analysis would take into account a restricted range of allowable angles.

The construction of an efficient alignment algorithm based on the distance $\mathcal{D}$ as a cost function is an important future goal, and could have important future implications for structure prediction and biomolecular folding dynamics. For our purposes here we chose the approximate metrics MRSD and RMSD. For the β-hairpin, the best-aligned MRSD structure was globally different than the best-aligned RMSD structure. The distance from a straight line to an idealized β-hairpin structure was slightly less when the structures were aligned by minimizing MRSD than for RMSD. However. the situation was reversed for the real β-hairpin structure, with the RMSD-aligned structures having a smaller distance by $\sim$5%.

The noncrossing transformation raises interesting questions about the validity of structural comparison metrics when polymer noncrossing is important. The RMSD and MRSD were both quite small for the conformations we considered, comparable to the α-helix distances. However, the actual distance for a physically realizable transformation was large—larger than the distances in β-sheet transformations.

The solution we found for the case of noncrossing was extremal and minimal, at least locally. However, there is no guarantee that this is the globally minimal transformation—some preliminary results for small numbers of links indicate there can be shorter pathways in some instances (A. R. Mohazab and S. S. Plotkin, unpublished data). However, the difference in distances between ground-state and excited-state transformations involves rotations of links and so is nonextensive: in the limit of large numbers of links, the discrepancies go to zero (43).

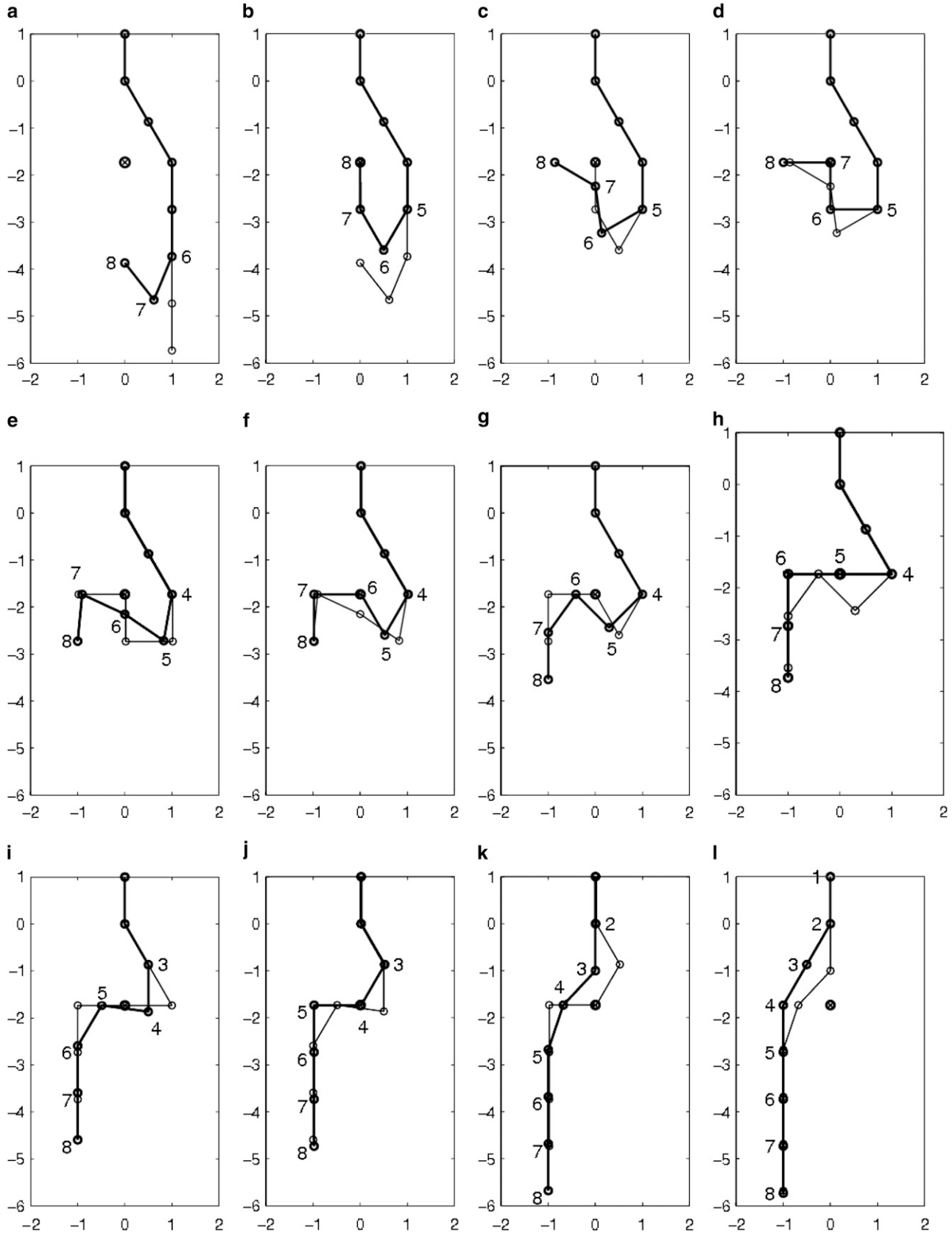

FIGURE 7 Various steps in a minimal pathway obeying noncrossing. Two conformations are drawn for each step. By convention, we number residues in the conformation that is leading in the transformation. (See text for a description of the transformation.)

Noncrossing constraints introduce a mechanistic aspect to the folding process. A folding mechanism consists of a specific sequence of events, or pathway. In the context of our problem the chain had to cross over the obstacle before translating to its final position. In practice the chain can go up and over the top or bottom of the obstacle, or cross over it in different places with varying likelihood, so strictly speaking there are many pathways and we have just investigated the minimal distance pathway here. Nevertheless, such constraints can further restrict the entropic bottleneck (49) governing folding rates.

The physics of noncrossing is certainly important for knotted proteins, and the generalized distance may be useful as an order parameter for these proteins, whereas other structural comparison parameters would be flawed. The noncrossing constraints in a knotted protein slow its kinetics (50,51), and lead to different molecular evolutionary pressures for fast and reliable folding (52–54).

For a simple stochastic process such as the one-dimensional diffusion of a point particle on a flat potential between two absorbing barriers, the splitting or commitment probability $p_F = \mathcal{D}/\mathcal{D}_{TOT}$, giving a correlation $\langle \mathcal{D} p_F \rangle = 1$. The presence of such a correlation between distance and commitment probability for simple examples provides encouragement to investigate whether or not one would find a significant correlation for the more complex problem of protein folding, in particular when the presence of noncrossing constraints for configurational diffusion has been accounted for. In the above discussion, p_F has tacitly been written in terms of $\mathcal{D}$ rather than the reverse. This underscores the conceptual importance of geometric order parameters in understanding the progress of a reaction.

An emergent simplicity in protein folding has been the conclusion that native topology determines the major features of the free energy landscape for a protein, and consequently a protein's folding rate and mechanism (46). The distance $\mathcal{D}$ between disordered or partly disordered protein structures and the native structure may capture the evolution of topology during the folding process more accurately than many other order parameters proposed to characterize the folding kinetics and mechanisms of proteins: a full systematic comparison remains a problem for future research.

Useful order parameters have simple geometric interpretations. Here we have shown that in principle one can compute the distance that would have to be traveled to connect two arbitrary biopolymer structures, a simple geometric quantity that can include noncrossing constraints, as well as properties such as restricted allowable angles or chain stiffness. The problem of finding a minimal distance pathway for a biomolecule is now an algorithmic problem rather than a conceptual one. In the long run, it is feasible that the analysis of chemical reactions involving large numbers of degrees of freedom might benefit from order parameters similar to the one we studied here, which are capable of accounting for the structural complexities inherent in large molecules.

APPENDIX

Point particle

The extremal trajectories of beads or links subject to steric excluded volume is a variational problem in the presence of an inequality constraint. A bead can be outside a given region but not inside it, or must travel from point A to point B while avoiding an intervening volume.

Variational problems subject to inequality constraints arose historically in the theory of optimal control (55–58). In our context we illustrate the idea with a simple example of a point particle moving from A to B but with the constraint that the point and resulting trajectory must lie outside an infinite cylinder of radius a, $r \geq a$ in Fig. 8.

The distance traveled by the point is written as

$$\mathcal{D}[\mathbf{r}] = \int_0^T dt\, F(\dot{\mathbf{r}}, \lambda, \epsilon), \tag{8a}$$

where

$$F(\dot{\mathbf{r}}, \lambda, \epsilon) = \int_0^T dt \left(\sqrt{\dot{\mathbf{r}}^2} + \lambda (a - |\mathbf{r}| + \epsilon^2) \right). \tag{8b}$$

The second term in the integrand embodies the inequality constraint $a - |\mathbf{r}| \leq 0$. The value λ is the Lagrange multiplier enforcing the constraint, and the quantity ϵ^2 may be thought of as an ''excess parameter'' whose significance will soon become clear.

Let a vector $X = (\mathbf{r}, \lambda, \epsilon)$ represent all the unknowns in the problem. The Euler-Lagrange (EL) equations are then

$$\frac{d}{dt} F_{\dot{X}} = F_X, \tag{9}$$

with the convention $F_X \equiv \partial F/\partial X$. The EL equations are

$$a - r + \epsilon^2 = 0, \tag{10a}$$

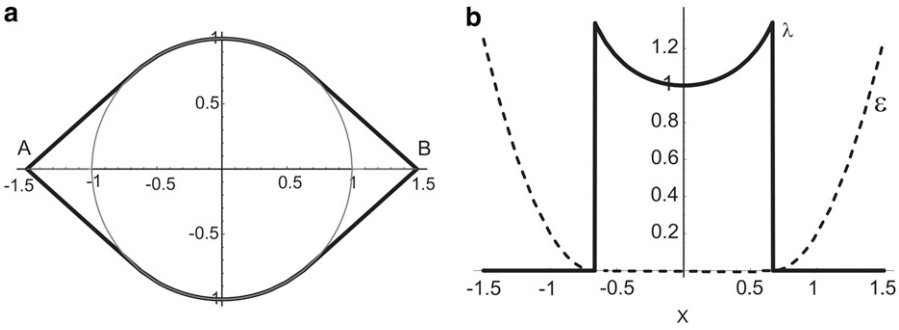

FIGURE 8 (*a*) Extremal trajectories for an inequality constraint problem. In this case, a path that is a minimal distance from point A at $(x_A, y_A) = (-1.5, 0)$ to point B at $(x_B, y_B) = (+1.5, 0)$ is sought subject to the constraint that the path must remain outside a circle of unit radius. Both positive and negative solutions are shown. (*b*) Lagrange multiplier λ and excess parameter ϵ for the above problem. If $\epsilon \neq 0, \lambda = 0$, and if $\epsilon = 0, \lambda \neq 0$.

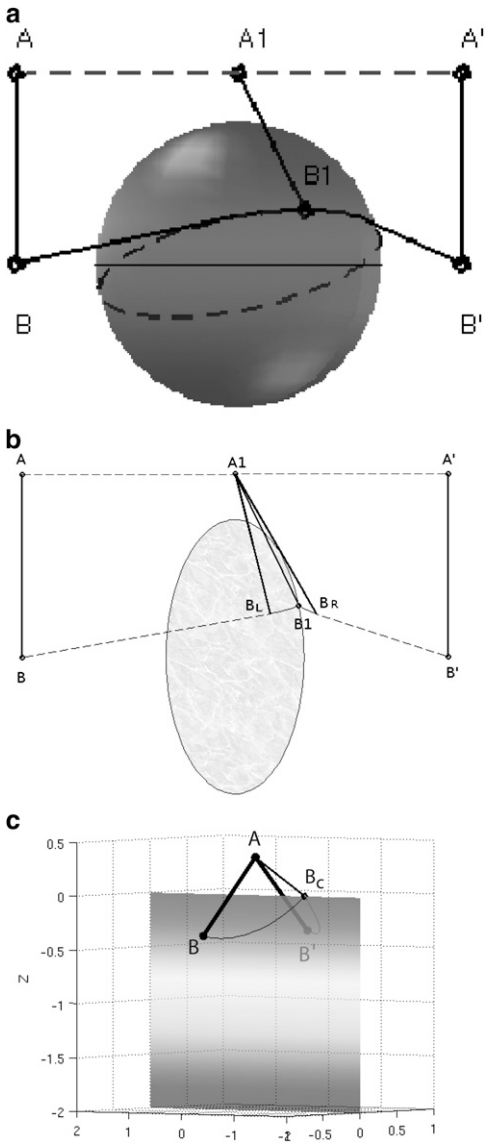

FIGURE 9 (*a*) Extremal trajectory for a one-link transformation subject to inequality constraints. The link moves from configuration AB to $A'B'$ in the presence of an obstructing sphere. The link length $\overline{AB}$ is conserved during this process. The distance traveled by the end-points A and B of the link is minimized by the transformation shown, which involves straight-line motion of A to A', and straight-line motion of B along a trajectory tangent to the sphere. Point B traces out a great circle on the surface of the sphere before continuing to B' on another trajectory tangent to the sphere. (*b*) When the sphere in panel *a* is compressed to form a two-dimensional disk of the same radius, the minimal transformation takes the form shown, with a discontinuity in the trajectory of B at point $B1$. Moreover, the piecewise solution must still retain rotations and is not purely piecewise straight lines. (*c*) Transformation from AB to AB', in the presence of an intervening infinite strip. The minimal transformation consists of two piecewise rotations with a corner violation between them: the link rotates from B to B_c, then from B_c to B'.

$$\epsilon\lambda = 0, \tag{10b}$$

$$\dot{\hat{\mathbf{v}}} = -\lambda\hat{\mathbf{r}}. \tag{10c}$$

In addition to the EL equations, transversality or corner conditions must hold for the trajectory to be extremal (59). These demand that

$$F_{\dot{\mathbf{r}}}(t^-) = F_{\dot{\mathbf{r}}}(t^+) \tag{11a}$$

and

$$F - \dot{\mathbf{r}}\cdot F_{\dot{\mathbf{r}}}\big|_{t^-} = F - \dot{\mathbf{r}}\cdot F_{\dot{\mathbf{r}}}\big|_{t^+}, \tag{11b}$$

where $t^{\pm} = \lim_{\epsilon\to 0}(t\pm\epsilon)$. In this parameterization ($\mathbf{r}$ in terms of time), Eq. 11b gives no new information, and Eq. 11a demands that

$$\hat{\mathbf{v}}(t^-) = \hat{\mathbf{v}}(t^+). \tag{12}$$

To solve these equations, first note that from Eq. 10a, if $r > a$, the excess parameter is $\epsilon^2 > 0$. Then from Eq. 10b, the Lagrange multiplier is $\lambda = 0$. Then from Eq. 10c, $\dot{\hat{\mathbf{v}}} = 0$ and the particle moves in a straight line. The particle moves in a straight line until a point where it touches the cylinder. Equation 12 demands that the straight line must be tangent to the cylinder, otherwise we would have a corner at that point. Once on the cylinder, $r = a$ and so $\epsilon^2 = 0$. The quantity $\dot{\hat{\mathbf{v}}}$ is determined kinematically by the trajectory which follows the boundary condition, here the surface of the cylinder at $r = a$. This then determines $\lambda(t) = |\dot{\hat{\mathbf{v}}}|$.

This gives the piecewise trajectory in Fig. 8 *a*. Both positive and negative solutions are shown. For this extremal trajectory, the Lagrange multiplier and excess parameter can be found straightforwardly, for example as functions of x (Fig. 8 *b*). In particular $\lambda = 1/y(x)$ on the cylinder, zero otherwise.

If the obstructing object is no longer a cylinder of circular cross section, but we compress the x axis of the cylinder so that it is an ellipsoid, then in the limit that the minor axis (the x axis of the ellipsoid) $\to 0$, the obstructing object becomes a flat strip (or line in cross section). Then the extremal trajectory consists of two straight-line pieces with an apparent corner between them, due to the discontinuity at the surface of the excluded boundary.

One link

The above solution can be generalized to the case of a single link undergoing a transformation from one side of a sphere to the other side. For the initial conditions in Fig. 9 *a*, the solution consists of one bead on the link moving in straight-line motion, and the other following a piecewise trajectory consisting of straight-line motion, a great circle geodesic, and finally straight-line motion again.

When one axis of the sphere is compressed so that the sphere becomes a disk, the minimal-distance solution acquires a discontinuity or cusp (Fig. 9 *b*). This means that minimal-distance transformations can violate corner conditions if the inequality constraints are themselves discontinuous or more precisely nonsmooth. The extremal transformation of the link $\overline{AB}$ in Fig. 9 *b* involves a straight-line translation of A to $A1$, while point B translates to B_L. Then point B rotates to point $B1$ on the surface of the disk, where it experiences a corner as per the above discussion. It subsequently rotates again to B_R, then $A1$ and B_R translate together in straight lines to points A' and B', respectively.

As another example, consider the initial conditions in Fig. 9 *c*, which involves the problem of one link transforming in the presence of an infinite strip. This situation has applications to the problem of chain noncrossing discussed in the text. The minimal transformation consists of two piecewise rotations of B with a corner between them, at position B_c.

Regarding the MRSD-aligned transformation

Implementing the transformation for the MRSD-aligned structures of the β-hairpin involves a somewhat subtle transformation for one of the links (the link denoted by the *small arrow* in Fig. 5 *i*). A representation of the transformation is shown in Fig. 10. This was the only instance of this kind of transformation we had found for the various configurations that we studied.

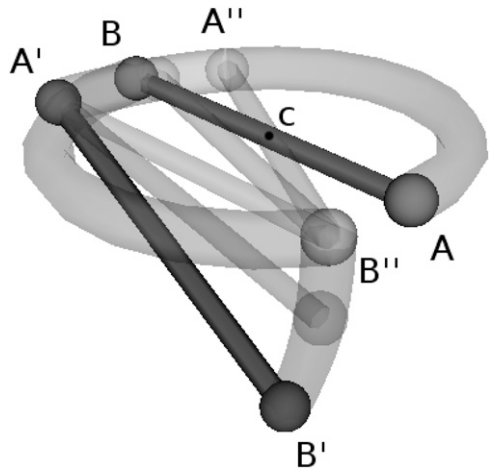

FIGURE 10 An uncommon one-link transformation that was nevertheless found for one of the links in the MRSD-aligned β-hairpin transformation in Fig. 5 *i* (link denoted by *small arrow*). The transformation obeys EL Eqs. 5a–5c and the corner conditions in Eq. 6. The initial and final conditions in this case are links AB and $A'B'$, respectively. Note the link has a direction—so, for example, A must transform to A' and not B'. The transformation proceeds by rotating link AB about point c to configuration $A''B''$. Bead A'' then rotates about B'' to position A', and B'' rotates about A' to position B'.

We thank C. Clementi and E. Abrahamsson for helpful discussions.

S.S.P. gratefully acknowledges support from the Natural Sciences and Engineering Research Council and the A. P. Sloan Foundation.

REFERENCES

1. Wolynes, P. G. 1997. Folding funnels and energy landscapes of larger proteins in the capillarity approximation. *Proc. Natl. Acad. Sci. USA.* 94:6170–6175.
2. Klimov, D. K., and D. Thirumalai. 1998. Lattice models for proteins reveal multiple folding nuclei for nucleation-collapse mechanism. *J. Mol. Biol.* 282:471–492.
3. Nymeyer, H., N. D. Socci, and J. N. Onuchic. 2000. Landscape approaches for determining the ensemble of folding transition states: success and failure hinge on the degree of minimal frustration. *Proc. Natl. Acad. Sci. USA.* 97:634–639.
4. Ozkan, S. B., K. A. Dill, and I. Bahar. 2003. Computing the transition state populations in simple protein models. *Biopolymers.* 68:35–46.
5. Fersht, A. R. 2004. Relationship of Leffler (Bronsted) α-values and protein folding ϕ-values to position of transition-state structures on reaction coordinates. *Proc. Natl. Acad. Sci. USA.* 101:14338–14342.
6. Bolhuis, P. G., D. Chandler, C. Dellago, and P. L. Geissler. 2002. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53:291–318.
7. Ejtehadi, M. R., S. P. Avall, and S. S. Plotkin. 2004. Three-body interactions improve the prediction of rate and mechanism in protein folding models. *Proc. Natl. Acad. Sci. USA.* 101:15088–15093.
8. Oztop, B., M. R. Ejtehadi, and S. S. Plotkin. 2004. Protein folding rates correlate with heterogeneity of folding mechanism. *Phys. Rev. Lett.* 93:208105.
9. Berezhkovskii, A., and A. Szabo. 2005. One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions. *J. Chem. Phys.* 122:014503.
10. Ding, F., W. Guo, N. V. Dokholyan, E. I. Shakhnovich, and J.-E. Shea. 2005. Reconstruction of the Src-SH3 protein domain transition state ensemble using multiscale molecular dynamics simulations. *J. Mol. Biol.* 350:1035–1050.
11. Weikl, T. R., and K. A. Dill. 2007. Transition-states in protein folding kinetics: the structural interpretation of ϕ-values. *J. Mol. Biol.* 365: 1578–1586.
12. Onsager, L. 1938. Initial recombination of ions. *Phys. Rev.* 54:554–557.
13. García, A. E. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699.
14. Chan, H. S., and K. A. Dill. 1994. Transition states and folding dynamics of proteins and heteropolymers. *J. Chem. Phys.* 100:9238–9257.
15. Plotkin, S. S., and P. G. Wolynes. 1998. Non-Markovian configurational diffusion and reaction coordinates in protein folding. *Phys. Rev. Lett.* 80:5015–5018.
16. Du, R., V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich. 1998. On the transition coordinate for protein folding. *J. Chem. Phys.* 108:334–350.
17. Hummer, G., A. E. García, and S. Garde. 2001. Helix nucleation kinetics from molecular simulations in explicit solvent. *Proteins.* 42:77–84.
18. Baumketner, A., J.-E. Shea, and Y. Hiwatari. 2004. Improved theoretical description of protein folding kinetics from rotations in the phase space of relevant order parameters. *J. Chem. Phys.* 121:1114–1120.
19. Ma, A., and A. R. Dinner. 2005. Automatic method for identifying reaction coordinates in complex systems. *J. Phys. Chem. B.* 109:6769–6779.
20. Best, R. B., and G. Hummer. 2005. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. USA.* 102:6732–6737.
21. Das, P., M. Moll, H. Stamati, L. E. Kavraki, and C. Clementi. 2006. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc. Natl. Acad. Sci. USA.* 103:9885–9890.
22. Cho, S. S., Y. Levy, and P. G. Wolynes. 2006. P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. USA.* 103:586–591.
23. Dokholyan, N. V., L. Li, F. Ding, and E. I. Shakhnovich. 2002. Topological determinants of protein folding. *Proc. Natl. Acad. Sci. USA.* 99:8637–8641.
24. Chavez, L. L., J. N. Onuchic, and C. Clementi. 2004. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* 126:8426–8432.
25. Wang, J., K. Zhang, H. Lu, and E. Wang. 2005. Quantifying kinetic paths of protein folding. *Biophys. J.* 89:1612–1620.
26. Wang, J., K. Zhang, H. Y. Lu, and E. K. Wang. 2006. Dominant kinetic paths on biomolecular binding-folding energy landscape. *Phys. Rev. Lett.* 96:168101.
27. Sega, M., P. Faccioli, F. Pederiva, G. Garberoglio, and H. Orland. 2007. Quantitative protein dynamics from dominant folding pathways. *Phys. Rev. Lett.* 99:118102.
28. Güner, U., Y. Arkun, and B. Erman. 2006. Optimum folding pathways of proteins: their determination and properties. *J. Chem. Phys.* 124:134911.
29. Chang, I., M. Cieplak, J. R. Banavar, and A. Maritan. 2004. What can one learn from experiments about the elusive transition state? *Protein Sci.* 13:2446–2457.
30. Plotkin, S. S., and J. N. Onuchic. 2002. Structural and energetic heterogeneity in protein folding. I: Theory. *J. Chem. Phys.* 116:5263–5283.
31. Wind, A. F., J. P. Kemp, A. V. Ermoshkin, and J. Z. Y. Chen. 2002. Structural and folding properties of a lattice prion model. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 66:031909.
32. Cheung, M. S., A. E. Garcia, and J. N. Onuchic. 2002. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA.* 99:685–690.
33. Karanicolas, J., and C. L. Brooks III. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11:2351–2361.

34. Snow, C., H. Nguyen, V. Pande, and M. Gruebele. 2002. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*. 420:102–106.

35. Simmerling, C., B. Strockbine, and A. Roitberg. 2002. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc*. 124:11258–11259.

36. Garcia, A. E., and J. N. Onuchic. 2003. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. USA*. 100:13898–13903.

37. Hummer, G., A. E. García, and S. Garde. 2000. Conformational diffusion and helix formation kinetics. *Phys. Rev. Lett*. 85:2637–2640.

38. Veitshans, T., D. Klimov, and D. Thirumalai. 1996. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Design*. 2:1–22.

39. Baumketner, A., and Y. Hiwatari. 2002. Diffusive dynamics of protein folding studied by molecular dynamics simulations of an off-lattice model. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 66:011905.

40. Cheung, M. S., and D. Thirumalai. 2006. Nanopore-protein interactions dramatically alter stability and yield of the native state in restricted spaces. *J. Mol. Biol*. 357:632–643.

41. Portman, J. J., S. Takada, and P. G. Wolynes. 2001. Microscopic theory of protein folding rates. I. Fine structure of the free energy profile and folding routes from a variational approach. *J. Chem. Phys*. 114:5069–5081.

42. Portman, J. J., S. Takada, and P. G. Wolynes. 2001. Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. *J. Chem. Phys*. 114:5082–5096.

43. Plotkin, S. S. 2007. Generalization of distance to higher dimensional objects. *Proc. Natl. Acad. Sci. USA*. 104:14899–14904.

44. Mohazab, A. R., and S. S. Plotkin. 2008. Minimal distance transformations between links and polymers: principles and examples. *J. Phys. Cond. Mat*. 20:244133.

45. Mines, G. A., T. Pascher, S. C. Lee, J. R. Winkler, and H. B. Gray. 1996. Cytochrome *c* folding triggered by electron transfer. *Chem. Biol*. 3:491–497.

46. Baker, D. 2000. A surprising simplicity to protein folding. *Nature*. 405:39–42.

47. Gouaux, J. E., and W. N. Lipscomb. 1990. Crystal structures of phosphonoacetamide ligated T and phosphonoacetamide and malonate ligated R states of aspartate carbamoyltransferase at 2.8-Å resolution and neutral pH. *Biochemistry*. 29:389–402.

48. Ozkan, S. B., G. A. Wu, J. D. Chodera, and K. A. Dill. 2007. Protein folding by zipping and assembly. *Proc. Natl. Acad. Sci. USA*. 104: 11987–11992.

49. Wolynes, P. G., J. N. Onuchic, and D. Thirumalai. 1995. Navigating the folding routes. *Science*. 267:1619–1620.

50. Finkelstein, A. V., and A. Y. Badretdinov. 1997. Influence of chain knotting on rate of folding. *Fold. Design*. 3:67–68.

51. Taylor, W. R. 2000. A deeply knotted protein structure and how it might fold. *Nature*. 06:916–919.

52. Mallam, A. L., and S. E. Jackson. 2006. Probing nature's knots: the folding pathway of a knotted homodimeric protein. *J. Mol. Biol*. 359: 1420–1436.

53. Virnau, P., L. A. Mirny, and M. Kardar. 2006. Intricate knots in proteins: function and evolution. *PLoS Comput. Biol*. 2:1074–1079.

54. Wallin, S., K. B. Zeldovich, and E. I. Shakhnovich. 2007. The folding mechanics of a knotted protein. *J. Mol. Biol*. 368:884–893.

55. Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. 1962. The Mathematical Theory of Optimal Processes. Wiley Interscience, New York and London.

56. Cass, D. 1965. Optimum growth in an aggregative model of capital accumulation. *Rev. Econ. Stud*. 32:233–240.

57. Gregory, J., and C. Lin. 1994. An unconstrained calculus of variations formulation for generalized optimal control problems and for the constrained problem of Bolza. *J. Math. Anal. Appl*. 187:826–841.

58. Gregory, J., and C. Lin. 2007. Constrained Optimization in the Calculus of Variations and Optimal Control Theory, 1st Ed. Springer, New York.

59. Gelfand, I. M., and S. V. Fomin. 2000. Calculus of Variations. Dover, Mineola, NY.