Measuring the average shape of transition paths during the folding of a single biological molecule

Noel Q. Hoffer[a], Krishna Neupane[a], Andrew G. T. Pyo[a], and Michael T. Woodside[a,1]

[a]Department of Physics, University of Alberta, Edmonton, AB T6G 2E1, Canada

Edited by William A. Eaton, National Institutes of Health, Bethesda, MD, and approved March 12, 2019 (received for review September 25, 2018)

Transition paths represent the parts of a reaction where the energy barrier separating products and reactants is crossed. They are essential to understanding reaction mechanisms, yet many of their properties remain unstudied. Here, we report measurements of the average shape of transition paths, studying the folding of DNA hairpins as a model system for folding reactions. Individual transition paths were detected in the folding trajectories of hairpins with different sequences held under tension in optical tweezers, and path shapes were computed by averaging all transitions in the time domain, $\langle t(x)\rangle$, or by averaging transitions of a given duration in the extension domain, $\langle x(t|\tau)\rangle_\tau$. Whereas $\langle t(x)\rangle$ was close to straight, with only a subtle curvature, $\langle x(t|\tau)\rangle_\tau$ had more pronounced curvature that fit well to theoretical expectations for the dominant transition path, returning diffusion coefficients similar to values obtained previously from independent methods. Simulations suggested that $\langle t(x)\rangle$ provided a less reliable representation of the path shape than $\langle x(t|\tau)\rangle_\tau$, because it was far more sensitive to the effects of coupling the molecule to the experimental force probe. Intriguingly, the path shape variance was larger for some hairpins than others, indicating sequence-dependent changes in the diversity of transition paths reflective of differences in the character of the energy barriers, such as the width of the barrier saddle-point or the presence of parallel paths through multiple barriers between the folded and unfolded states. These studies of average path shapes point the way forward for probing the rich information contained in path shape fluctuations.

energy landscapes | diffusive reactions | DNA hairpins | optical tweezers

Transition paths involve those segments of a reaction during which the energy barrier between reactants and products is crossed (1, 2). They represent the most interesting part of any reaction because they exclude the nonproductive fluctuations, focusing only on the productive portions of the trajectories (Fig. 1). In particular, the high-energy states occupied along the transition paths dominate the reaction kinetics and effectively encode the reaction mechanism. Transition paths are especially interesting for understanding the folding of biological molecules like proteins and nucleic acids, because of the variety and complexity of possible mechanisms. They are technically challenging to measure in folding reactions, however, because they can only be observed in single molecules and have a very brief duration. As a result, it has only recently become possible to measure transition path properties directly (3). Work to date using fluorescence and force spectroscopy has probed properties such as the average transition path time for proteins and nucleic acids (4–8), the variations in transit times for individual transitions (9, 10), the occupancy statistics within transition paths (11, 12), the distribution of velocities along the transition paths (13), and the agreement between experiment and theory for the properties measured to date (13–15).

These studies have yielded important insight into the microscopic motions involved in folding, but they have been limited to characterizing the average values and distributions of properties like transition path times and velocities. Such averages and distributions, while reflecting key aspects of the folding, nevertheless do not retain any information about the sequence of events taking place during the transitions. The ability to study sequential relationships between the local motions that constitute the folding reaction promises to unlock a wealth of information for characterizing folding mechanisms by directly observing the statistical ensemble of pathways populated during a folding reaction. Such direct observation has not yet been possible in experiments, restricting the study of pathway ensembles to simulations (16–23). However, the desired sequential information can in principle be accessed through the shapes of the transition paths, which capture the time spent within each part of the reaction coordinate and how it varies as the energy barrier is crossed.

Recent theoretical work has begun to explore the properties of transition path shapes, focusing on characterizing the average shape. Kim and Netz (24) considered the shape defined by $\langle t(x)\rangle$, the average of the transition paths in the time domain as a function of position along the reaction coordinate (which effectively reflects the distribution of times required to reach each point x along the transition), deriving expressions for the average path shapes expected for different types of barrier potential. Makarov (25) and Cossio et al. (26) took a somewhat different approach, focusing instead on the average shape of the dominant or most probable path, predicting the path shape expected for transitions crossing a harmonic barrier. This work showed that the transition path shapes are more sensitive to the shape of the underlying energy barrier than are properties like the transition time, suggesting that path shapes should indeed be a powerful tool for probing the microscopic details of folding.

These theoretical predictions have not been compared with experimental observations, however, because measurements of transition path shapes have not yet been reported. As a first step toward experimental characterization of transition path shapes,

Significance

Transition path shapes encode key information about folding mechanisms because they represent the sequence of microscopic motions followed while traversing the critical transition states within the energy barrier separating folded and unfolded states. However, they are technically challenging to measure and have not been observed experimentally. We present measurements of transition path shapes, determining the average shapes for the folding of DNA hairpins. We find that time domain averages are sensitive to instrumental artifacts, but extension domain averages match the dominant path shape predicted theoretically for diffusion over a 1D landscape. Path shape variances reveal sequence-dependent changes in the diversity of transition paths, suggesting that multiple parallel pathways through the transition states can be detected simultaneously in a single molecule.

Author contributions: N.Q.H. and M.T.W. designed research; N.Q.H., K.N., and A.G.T.P. performed research; N.Q.H. analyzed data; and N.Q.H. and M.T.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: michael.woodside@ualberta.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1816602116/-/DCSupplemental.

Published online April 5, 2019.

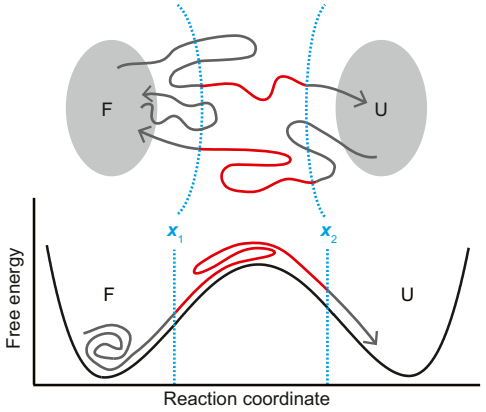

Fig. 1. Transition paths. Transition paths (red) represent the productive parts of a folding trajectory (gray) where the free energy barrier separating the folded (F) and unfolded (U) states is crossed. The boundaries of the barrier region are labeled x_1 and x_2 (cyan).

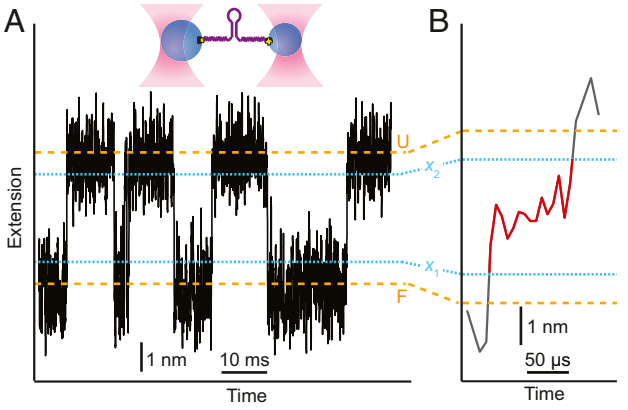

Fig. 2. Transition path measurements of DNA hairpins. (*A*) The end-to-end extension of a single hairpin held under tension by optical tweezers fluctuates between the folded (F) and unfolded (U) states. (*Upper Inset*) Cartoon of measurement. The hairpin is attached via dsDNA handles to polystyrene beads held in laser traps. (*B*) A transition path crosses the barrier region from x_1 to x_2 (cyan) without returning (gray, hairpin folding trajectory; red, transition path).

here we have used single-molecule force spectroscopy measurements of DNA hairpins held under tension in high-resolution optical tweezers as a model system to determine the average shapes of the transition paths computed by averaging in the time domain and in the extension domain. The transition path shapes were obscured in time domain averages by the effects of the experimental force probes, but for each hairpin the extension domain average matched the shape of the dominant transition path predicted for diffusion over a harmonic barrier, yielding diffusion coefficients similar to previous results. We also explored the fluctuations in the path shapes, finding evidence that more than one type of transition path may be present for some hairpins.

Results

To measure transition paths, we attached single DNA hairpins to polystyrene beads held in optical traps, using kilobase-long double-stranded DNA to link the hairpins to the beads (Fig. 2*A*, *Inset*). The hairpin constructs, which had different stem sequences to produce different energy landscapes (hairpins 30R50/T4, 20R100/T4, 20R55/T4, and 20R25/T4 from ref. 27 as well as 20TS06/T4 from ref. 28), were held under tension at constant trap separation with a force near $F_{1/2}$, the value at which the folded and unfolded states were equally probable, while measuring the end-to-end extension as the hairpin structure fluctuated in equilibrium between the two states (Fig. 2*A*). High trap stiffness ensured a time resolution of ∼6–9 μs (9). We identified transition paths within the extension trajectories as those parts of a trajectory passing between two boundaries, x_1 and x_2, demarking the barrier region (Fig. 2*B*). For each type of hairpin, ∼12,600–66,000 transitions were measured.

Isolating a representative set of individual transitions from within the full folding trajectory for a single hairpin and aligning the transitions to start at the same time so as to characterize the transition path shapes (Fig. 3), a large variety of path shapes was seen for both unfolding (Fig. 3*A*, gray) and refolding (Fig. 3*B*, gray). This wide range of shapes represents many different patterns of motion for the hairpin as it crosses the energy barrier. We first found the average shape of the transitions for each hairpin by expressing the transition path trajectories as $t(x)$, aligning them all on the boundaries of the barrier region (x_1 and x_2), and then averaging in the time domain to obtain $\langle t(x)\rangle$. Because many trajectories involved recrossing events where the hairpin extension turned back on itself before completing the transition (13), we took the average of all of the times at which a given x value was crossed in a particular transition path before averaging the values across all different paths, a procedure shown to avoid artifactual overweighting of transitions that include recrossing events (26). The averages for the unfolding (Fig. 3*A*, black) and refolding (Fig. 3*B*, red) transition paths for hairpin 30R50/T4, shown overlaid on 30 individual path trajectories (Fig. 3 *A* and *B*, gray), were found to be time-reversal symmetric (Fig. 3*C*), as expected from the microscopic reversibility of transition paths (25). Similar results were found for the other hairpins (*SI Appendix*, Fig. S1). Despite the wide variation in the individual path shapes, the average shapes computed from $\langle t(x)\rangle$ were all very similar and quite simple: almost straight lines, but with a very slight sigmoidal curvature where the path shape curved up on the

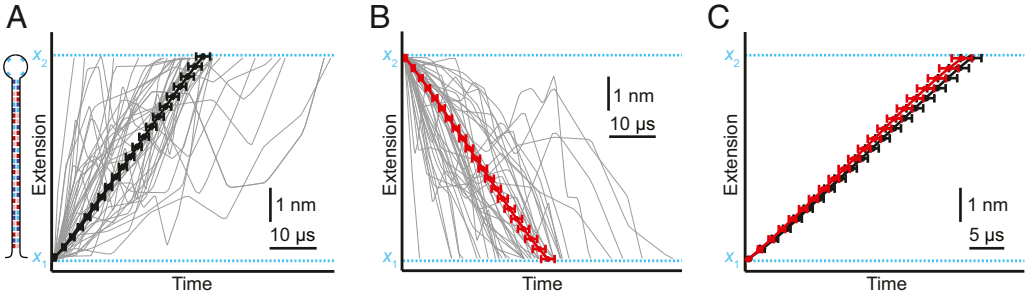

Fig. 3. Time domain average transition path shape $\langle t(x)\rangle$ for hairpin 30R50/T4. (*A*) The path shape for unfolding averaged in the time domain (black), shown along with 30 individual unfolding transitions (gray), is almost straight but with a slight sigmoidal curvature. (*B*) The average path shape $\langle t(x)\rangle$ for refolding transitions (red), shown with 30 individual refolding transitions (gray), is the same but time reversed. (*C*) The time-reversed path shape for refolding transitions (red) matches the shape for unfolding transitions (black). Error bars represent SEM.

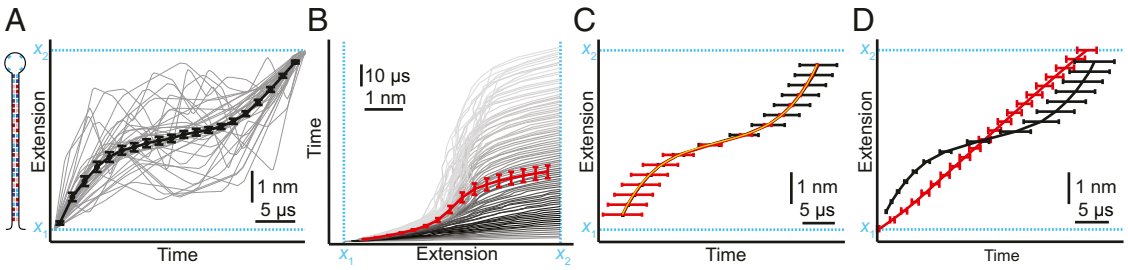

Fig. 4. Extension domain average transition path shape $\langle x(t|\tau)\rangle_\tau$ for hairpin 30R50/T4. (*A*) The path shape averaged in the extension domain for unfolding transitions with duration τ_{TP} (black), shown along with 30 individual unfolding transitions (gray). (*B*) The average path shape $\langle x(t|\tau)\rangle_\tau$ for unfolding (red) was computed by averaging the curves $\langle x(t|\tau_i)\rangle$ (gray) obtained from transition paths with duration τ_i, for all τ_i values. Grayscale coloring indicates the proportion of transitions with duration τ_i. (*C*) The path shapes $\langle x(t|\tau)\rangle_\tau$ for unfolding (black) and refolding (red; time reversed) are time reversal symmetric and well fit by Eq. **1** (yellow). The fits for unfolding and refolding are visually indistinguishable. (*D*) The time domain average shape $\langle t(x)\rangle$ (red) differs noticeably from the extension domain average shape $\langle x(t|\tau)\rangle_\tau$ (black), with the latter having much more pronounced curvature. Error bars represent SEM.

side of the barrier region nearest the folded state and down on the side nearest the unfolded state.

The conceptual motivation for calculating path shapes by averaging in the time domain at each extension value, as above, is that every transition passes through the same range of extension values, from x_1 to x_2, making the simple average $\langle t(x)\rangle$ well-defined. In contrast, a simple average in the extension domain at each time value—that is, $\langle x(t)\rangle$—is ill-defined, because the range of t values varies widely between transitions (10, 29, 30). Experimentally, however, averaging in the extension domain is more natural, since x is the observed variable containing experimental noise that should be averaged, not t. To compute an extension domain average while accounting for the differences in the transition path times, we averaged all of the transitions having a given transition path time τ_i to obtain $\langle x(t)|\tau_i\rangle$, illustrated in Fig. 4*A* for hairpin 30R50/T4, and then averaged the resulting curves from all τ_i (Fig. 4*B*, gray) in the time domain, weighting by the relative probability for each τ_i value. The average shapes computed in this way (Fig. 4*B*, red), denoted $\langle x(t|\tau)\rangle_\tau$, were the same for unfolding (Fig. 4*C*, black) and refolding (Fig. 4*C*, red), again reflecting the expected time-reversal symmetry. They showed a pronounced curvature that differed from the shape of $\langle t(x)\rangle$: Whereas both were sigmoidal, the curvature was opposite in sign and noticeably higher for the extension domain average (Fig. 4*D*, black) than the time domain average (Fig. 4*D*, red). Similar results were observed for the other hairpins (*SI Appendix*, Fig. S2).

To understand why these two methods of computing the average shapes yielded different results, we explored Brownian-dynamics simulations of transitions paths. Simulating diffusive crossing of a one-dimensional (1D) harmonic barrier as described previously (13), we found that both $\langle t(x)\rangle$ and $\langle x(t|\tau)\rangle_\tau$ had similar shapes (Fig. 5*A*, red and black, respectively), in contrast to the experimental results. Furthermore, both averages closely matched the shape of the dominant transition path (Fig. 5*A*, blue) calculated analytically from the potential (25, 26, 31). However, previous work has shown that attaching a molecule to a large probe like a bead or cantilever via a compliant linker and monitoring the motions of the probe rather than those of the molecule, as is done in force spectroscopy measurements, adds noise to the measured extension and can alter the properties of the transition paths (26, 32–35). We therefore extended the simulations to include the effects of linking the molecule to a bead and recomputed $\langle t(x)\rangle$ and $\langle x(t|\tau)\rangle_\tau$. In contrast to $\langle x(t|\tau)\rangle_\tau$ (Fig. 5*B*, black), which was little changed from the 1D simulation, $\langle t(x)\rangle$ was distorted to become close to linear (Fig. 5*B*, red), analogous to what was observed experimentally. The simulations thus suggest that time domain averaging can lead to an unreliable picture of the average path shape owing to the effects of the experimental geometry, whereas averaging in the extension domain is more robust.

We next compared the average path shapes obtained from the experimental data to theory. Since simulations showed that $\langle x(t|\tau)\rangle_\tau$ provides a reasonable approximation of the analytical shape of the dominant transition path for harmonic barriers (Fig. 5*A*), and since the transition paths in hairpin folding are generally well-described by harmonic-barrier models (9, 11, 15), we fit the average shapes to the functional form of the dominant path shape for a harmonic barrier. In 1D with constant diffusion, this path shape is given by the following:

$$\langle x(t)\rangle_{DTP} = \frac{L\sinh(\beta\kappa D(t-\tau_{TP}/2))}{\sinh(\beta\kappa D\tau_{TP}/2)}, \qquad [1]$$

where $2L$ is the length of the transition path, κ is the curvature of the barrier, τ_{TP} is the average transition path time, and D is the diffusion coefficient for motion along the reaction coordinate (26). Fitting $\langle x(t|\tau)\rangle_\tau$ for hairpin 30R50/T4 to Eq. **1**, while fixing $2L = x_2 - x_1$, using the values of κ found from energy-landscape reconstructions (9), and treating D as a free parameter, we found good agreement with the observed shape (Fig. 4*C*, yellow). The values of D returned by the fits, $6 \pm 4 \times 10^5$ nm^2/s for both unfolding and refolding, agreed within error with the values in the range 2–5 $\times 10^5$ nm^2/s found previously for this hairpin by analyzing rates (6), transition path times (9), and transition path velocities (13). Repeating this analysis for each hairpin by fitting $\langle x(t|\tau)\rangle_\tau$ to Eq. **1** (*SI Appendix*, Fig. S2), we found similar results in each case (Table 1), with values of D that were consistent within error with previous results, although as for hairpin 30R50/T4 they were all systematically somewhat higher.

Finally, to obtain a measure of the variability in the transition path shapes, we calculated the variance in $t(x)$, $\sigma^2_{t(x)}$, for each of the hairpins as a function of the progression along the transition path (Fig. 6). The variance was very similar in the unfolding and refolding path shapes; hence the two were averaged for each hairpin. Not surprisingly, given the broad distribution of transit times (9), the variance in $t(x)$ increased superlinearly with increasing x. To our knowledge, there is not yet any theoretical analysis of path shape variance to which these results can be compared. However, comparison of $\sigma^2_{t(x)}$ between different hairpins revealed some noticeable differences: After traveling a

Table 1. Diffusion coefficients from fitting dominant path shapes

	D ($\times 10^5$ nm^2/s)		
Hairpin	Unfolding	Refolding	Average
30R50/T4	6 ± 4	6 ± 4	6 ± 4
20R100/T4	7 ± 4	7 ± 4	7 ± 4
20R55/T4	8 ± 4	8 ± 4	8 ± 4
20TS06/T4	4 ± 3	5 ± 3	4 ± 3
20R25/T4	7 ± 3	7 ± 3	7 ± 3

Errors represent SD from bootstrapping analysis of fits.

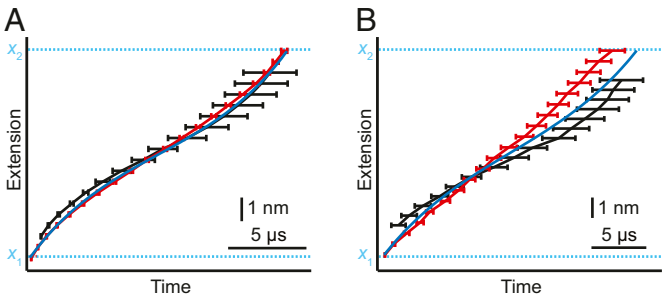

Fig. 5. Transition path shapes from Brownian-dynamics simulations. (*A*) Brownian-dynamics simulations of transitions over a 1D harmonic barrier show that both $\langle t(x)\rangle$ (red) and $\langle x(t|\tau)\rangle_\tau$ (black) match the dominant transition path shape computed analytically (blue). (*B*) Brownian-dynamics simulations of a bead coupled via a compliant linker to a molecule crossing the same barrier as in *A* show that the bead/linker effects cause $\langle t(x)\rangle$ to become more linear while leaving $\langle x(t|\tau)\rangle_\tau$ similar to the dominant transition path shape.

given distance x along a transition path, the variance in $t(x)$ was systematically lowest for hairpin 20R100/T4 and highest for hairpin 30R50/T4, whereas the remaining hairpins had variance between these two limits. We note that the absolute magnitudes of the observed variances are not intrinsically meaningful, because they depend systematically on the time resolution of the measurement, with many of the microscopic fluctuations being smoothed out by the instrument (13, 36). However, when comparing the variances for different measurements using the same instrument and conditions, as here, the instrumental effects are constant, and hence differences in the variance reflect meaningful differences in the variability of the transition paths sampled by different molecules.

Discussion

This work reveals some interesting features of transition path shapes and how to compute them. An important practical result is that the appealingly simple definition of the average path shape in terms of $\langle t(x)\rangle$, as proposed previously (24–26), is not reliable when applied to experimental data. Although it works well for pure 1D simulations, it is quite sensitive to experimental effects like the attachment of the molecule to force probes, which induce distortions that obscure the true transition path shape resulting from the underlying energy landscape by causing it to become more linear. This problem will likely apply to all measurement methods, not just force spectroscopy, because much of the distortion in $\langle t(x)\rangle$ appears to arise from averaging in the independent variable, t, instead of the dependent variable x that contains the measurement noise: When we generated an ensemble of simulated transition paths simply by adding Gaussian noise to the analytical path shape defined by Eq. **1**, we found that $\langle t(x)\rangle$ became more linear (*SI Appendix*, Fig. S3, red), similar to the effect seen in the force spectroscopy data and simulations. In contrast, the extension domain average $\langle x(t|\tau)\rangle$ was not affected (*SI Appendix*, Fig. S3, black), because the noise was correctly averaged in the measured variable x and the computed shape thus remained close to the original analytical path shape (*SI Appendix*, Fig. S3, blue).

Because $\langle t(x)\rangle$ is distorted by instrumental effects, it is unreliable for reporting on the properties of the barrier, even though simulations indicate that it should be quite sensitive to the barrier shape (25). However, there is an intriguing connection between the anomalously low barrier height implied by the flatter-than-expected shape of $\langle t(x)\rangle$ and previous work finding that transition path time distributions also implied barrier heights significantly lower than measured directly (10). Recent theoretical work has shown that distortions in the transition path time distribution implying anomalously low barriers can be induced by memory effects in the folding dynamics (37). Such memory effects would be expected to arise naturally in force spectroscopy measurements from the compliant linker coupling the molecule to the force probe, because of the finite response time for propagation of molecular motions through the linker to the probe. Hence the distortions of $\langle t(x)\rangle$ could also be viewed, at least in part, as a consequence of linker/probe-induced memory effects.

Turning to the extension domain average, the fact that $\langle x(t|\tau)\rangle_\tau$ matches the functional form of the dominant transition path shape suggests that this average provides a reasonable, heuristic approximation to the dominant path. Even though there is no formal equivalence between these two quantities, it is intuitively appealing that an appropriately constructed average of the path shapes should be able to approximate the dominant path shape. Indeed, such reasoning suggests an even simpler heuristic: approximating the dominant path not as the average of all paths but only of those close to the average transition path time, which are taken in this case as representative of the average behavior of the ensemble of curves. Calculating this restricted average, $\langle x(t|\tau_{TP})\rangle$ (*SI Appendix*, Fig. S4, black), we found that it was indeed very similar to the more general average $\langle x(t|\tau)\rangle_\tau$ computed previously (*SI Appendix*, Fig. S4, gray), and it, too, fit very well to Eq. **1** for the shape of the dominant transition path (*SI Appendix*, Fig. S4, red). The values for D obtained from the fits were very similar to those in Table 1 (*SI Appendix*, Table S1), again being close to the results obtained by other methods but with a systematic tendency to be a bit higher. Hence $\langle x(t|\tau_{TP})\rangle$ can be used as an approximation to the dominant path shape that is simpler to compute than $\langle x(t|\tau)\rangle_\tau$.

The reasonable agreement between the extension domain averages and expectations for a 1D harmonic barrier with constant D is perhaps not surprising, given that this same approximation has worked well for describing other transition path properties such as the average transition time, the distribution of transit times, and the average velocity at the barrier peak (9, 11, 13). However, even though the values of D implied by the fits to Eq. **1** are consistent within error with the values found from previous analyses of transition path properties in the same limit, they are systematically somewhat higher: in the range ~4–8 × 10^5 nm^2/s, compared with ~2–5 × 10^5 nm^2/s (9, 11, 13). Analysis of the simulations indicates that most of this discrepancy likely arises from the use of $\langle x(t|\tau)\rangle_\tau$ as an imperfect approximation for the dominant path shape, since the value of D from fitting $\langle x(t|\tau)\rangle_\tau$ to Eq. **1** was ~50% higher than the actual value imposed in the simulations. Other factors involving deviations from the assumptions underlying Eq. **1** may also contribute to the overestimate, however, including the presence of anharmonicity in the measured barrier profiles (9, 37), the possibility of position dependence in the diffusivity (34, 38), the presence of memory effects (36, 37), and the finite barrier size (14).

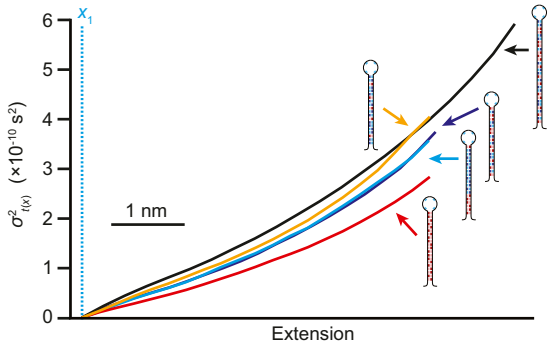

Fig. 6. Comparison of path shape variance for different hairpins. The variance in the average transition path shape, $\sigma^2_{t(x)}$, at each position along the reaction coordinate is similar for hairpins 20R25/T4, 20R55/T4, and 20TS06/T4, smallest for hairpin 20R100/T4, and largest for hairpin 30R50/T4. Larger variance suggests the presence of additional types of transition paths, providing evidence for multiple pathways.

Last, we consider the variability in path shapes. Naively, the fact that the average $\langle x(t|\tau)\rangle_\tau$ matches the shape predicted by Eq. **1** for the dominant transition path suggests that the observed paths primarily involve transitions over a single, 1D barrier that is largely harmonic, that is, the path variability reflects the diversity of shapes expected inherently from diffusive motion over this barrier. However, comparing the variance in average path shapes for the different hairpin sequences reveals that the situation must be more complex. Indeed, the differences in $\sigma^2_{t(x)}$ seen between the hairpins in Fig. 6 indicate that the diversity of transition paths occupied during the hairpin folding is sequence dependent: The transitions in some hairpins (e.g., 30R50/T4) display a broader range of path shapes, whereas those in others (e.g., 20R100/T4) occupy a more restricted range. These differences must reflect changes in the character of the energy barrier separating the folded and unfolded states. As one possibility, if the transition paths pass through only a single saddle-point in the landscape, then a broader saddle-point (reflecting, say, a larger transition state ensemble) would increase the variance in path shapes. Another possibility is that increased variance could reflect the presence of multiple, parallel paths through the transition states such that there are transition paths crossing through different saddle-points in the landscape. A multiplicity of microscopically distinct paths across the energy barrier is certainly expected from the statistical nature of the energy landscape picture of folding (39–41), and parallel pathways have been both observed directly in multistate folders (42–45) and inferred from the kinetics of two-state and downhill folders (46–49), but it has not previously been possible to quantify the extent of the diversity of transition paths in a folding reaction. Regardless of the precise origins of the differences in diversity of transition path shapes observed here, however, we note that simple 1D models of folding are not sufficient to account for such differences—despite the quantitative success of these models in explaining other properties like transition path times (9, 10), velocities (13), and occupancies (11, 12)—underlining the notion that transition path shapes are more sensitive to the microscopic details of the energy landscape than other such properties.

Conclusion

Here, we have presented measurements of transition path shapes in a folding reaction, computing different definitions of the average path shape for DNA hairpin folding measured with force spectroscopy and comparing them to theory. We found that averaging in the time domain led to distortions in the path shape arising from experimental noise and coupling of the hairpins to the force probe, making this definition of the path shape unreliable for application to experiments. However, averaging in the extension domain led to a shape matching the dominant path shape predicted theoretically for a 1D harmonic barrier, returning reasonable values for the conformational diffusion coefficient. Moreover, differences in the variance among the path shapes for different hairpins suggested that subpopulations of transition paths with different microscopic properties are present in these measurements. This work opens the door for future experimental and theoretical efforts to detect different populations of transition paths through their distinct characteristics and thereby probe the full range of microscopic behaviors present in biomolecular folding.

Materials and Methods

Sample Preparation. DNA hairpins attached to kilobase-long dsDNA handles were prepared as described previously (27). Briefly, one handle made via autosticky PCR with a biotin label at one end and a hairpin separated from the handle with abasic sites at the other end was ligated to a second handle labeled at one end with digoxigenin and at the other end with a single-stranded overhang complementary to an overhang on the 5′-end of the hairpin. Hairpin–handle constructs were incubated with 600-nm-diameter biotin-labeled polystyrene beads and 820-nm diameter anti-digoxigenin–labeled beads to form dumbbells, which were then diluted into the measurement buffer (50 mM Mops, pH 7.5, 200 mM KCl) containing an oxygen scavenger [8 mU/μL glucose oxidase, 20 mU/μL catalase, 0.01% (wt/vol) D-glucose] and inserted into a sample cell for measurement.

Transition Path Measurements. Folding trajectories were measured as described previously (9) by moving the traps apart at constant speed to ramp the force up to near $F_{1/2}$, and then holding the trap separation constant while the hairpin fluctuated between folded and unfolded states. The end-to-end extension of the construct was sampled at 125–1,000 kHz and filtered on-line at the Nyquist frequency. The stiffnesses of the two traps were kept high, 0.75–1.1 pN/nm in one trap and 0.56–0.66 pN/nm in the other, to maximize the response time of the instrument, which was measured previously as ∼6–9 μs (9). Under these measurement conditions, the kinetic artifacts affecting the observed rates and transition path properties that arise from the beads and handles (26, 32) were small (50).

Path Shape Analysis. The transition path shape averaged in the time domain, $\langle t(x)\rangle$, was determined from individual transitions extracted from extension trajectories, aligning the transitions at their starting points (set to $t = 0$). For each transition, the time at which the trajectory crossed a particular extension value was recorded; if the transition crossed a given extension multiple times, then the average value of the crossing times was taken (26). This procedure generated a single-valued path $t(x)$ for each transition, which was averaged at each x value over all of the transitions to obtain $\langle t(x)\rangle$, and effectively ensured that $\langle t(x)\rangle$ was symmetrized as discussed in refs. 25 and 26. Because fluctuations near the boundaries of the barrier region may cause the trajectory to prematurely cross the boundaries x_1 or x_2 before the end of the transition path, a fixed number of data points beyond the first boundary crossing were included in the averages, such that the average time to complete the transition was consistent with the values obtained previously.

The transition path shape averaged in the extension domain, $\langle x(t|\tau)\rangle_\tau$, was determined by extracting from the folding trajectories all of the transitions with transition times in a 1-μs band around a given value τ_i. The transition times for each folding and unfolding event were found from the extension trajectories as described previously (10). All of the transitions with duration τ_i were then aligned at their starting points (set to $t = 0$), and the values of x were averaged for each value of t to obtain $\langle x(t|\tau_i)\rangle$. This procedure was repeated for each transition time τ_i to generate a family of curves showing the path shapes for different τ_i when averaged in the extension domain. The shapes for different transition times were then combined to obtain the overall average, $\langle x(t|\tau)\rangle_\tau$, by inverting the shapes $\langle x(t|\tau_i)\rangle$ for every τ_i and averaging the time values for each point x. Defining $S_i(t) = \langle x(t|\tau_i)\rangle$, the overall average was then $\langle x(t|\tau)\rangle_\tau = \sum_i w_i S_i^{-1}(x)$, where the average is weighted by w_i, the normalized probability of finding a transition path of duration τ_i. When fitting the path shapes to Eq. **1**, the uncertainty in the diffusion coefficient was calculated as the SD of the fitting results from a bootstrapping analysis with 10,000 iterations.

Transition Path Simulations. One-dimensional Brownian-dynamics simulations of transition paths over a harmonic barrier were performed as described previously (13) using the method of Lu and Nolen (51), wherein the trajectories of transition paths are found by iteratively solving the Itô stochastic differential equation. For Brownian-dynamics simulations of a bead compliantly linked under tension to the molecule, the overdamped Langevin equations along the molecular extension x and the bead coordinate q were integrated as described previously (32, 52). Briefly, with the 2D potential of the bead-molecule system given by $G(x, q) = G_0(x) + 1/2k(q - x)^2$, for a molecular potential $G_0(x)$ and linker spring stiffness k at the applied force, the equations of motion are as follows:

$$x(t+\delta t) = x(t) - \delta t \beta D_x \partial G(x,q)/\partial x + \sqrt{2D_x \delta t} R_x(t)$$

and

$$q(t+\delta t) = q(t) - \delta t \beta D_q \partial G(x,q)/\partial q + \sqrt{2D_q \delta t} R_q(t),$$

where δt is the time step, β is the inverse thermal energy, D_x is the diffusion coefficient of the molecule, D_q is the diffusion coefficient of the bead, and $R_{x/q}$ is independent normally distributed random forces with zero mean and unity variance. Time steps for all simulations were 10 ns. All simulated trajectories were analyzed as done for the experimental trajectories.

To explore the effects of measurement noise on the analysis, noise drawn from a Gaussian distribution with $\sigma = 1$ nm (similar to experimental noise levels) was added randomly at each point to the ideal transition path shape described by Eq. **1**. This process was repeated 1,000 times to generate an

ensemble of "noisy" transition trajectories, which were then analyzed as done for the experimental transitions.

Variance Analysis. The variance $\sigma^2_{t(x)}$ was calculated at the same time as the average $\langle t(x)\rangle$. The results for unfolding and refolding transitions were nearly identical; hence the average of the two was computed at each point. Note that the variance in $\langle x(t|\tau)\rangle_\tau$ is not meaningful, as the way the average is computed means that it simply reflects the width of the distribution of transition path times.

Data Availability. Data are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS. We thank Attila Szabo and Dmitrii Makarov for helpful discussions. This work was supported by Natural Sciences and Engineering Research Council Canada and Alberta Innovates Technology Futures. M.T.W. acknowledges support from the John Simon Guggenheim Foundation.

1. Hummer G (2004) From transition paths to transition states and rate coefficients. *J Chem Phys* 120:516–523.
2. E W, Vanden-Eijnden E (2010) Transition-path theory and path-finding algorithms for the study of rare events. *Annu Rev Phys Chem* 61:391–420.
3. Chung HS, Eaton WA (2018) Protein folding transition path times from single molecule FRET. *Curr Opin Struct Biol* 48:30–39.
4. Chung HS, McHale K, Louis JM, Eaton WA (2012) Single-molecule fluorescence experiments determine protein folding transition path times. *Science* 335:981–984.
5. Chung HS, Eaton WA (2013) Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* 502:685–688.
6. Neupane K, et al. (2012) Transition path times for nucleic acid folding determined from energy-landscape analysis of single-molecule trajectories. *Phys Rev Lett* 109: 068102.
7. Truex K, Chung HS, Louis JM, Eaton WA (2015) Testing landscape theory for biomolecular processes with single molecule fluorescence spectroscopy. *Phys Rev Lett* 115:018101.
8. Yu H, et al. (2012) Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *Proc Natl Acad Sci USA* 109:14452–14457.
9. Neupane K, Wang F, Woodside MT (2017) Direct measurement of sequence-dependent transition path times and conformational diffusion in DNA duplex formation. *Proc Natl Acad Sci USA* 114:1329–1334.
10. Neupane K, et al. (2016) Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* 352:239–242.
11. Neupane K, Manuel AP, Woodside MT (2016) Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nat Phys* 12:700–703.
12. Neupane K, Manuel AP, Lambert J, Woodside MT (2015) Transition-path probability as a test of reaction-coordinate quality reveals DNA hairpin folding is a one-dimensional diffusive process. *J Phys Chem Lett* 6:1005–1010.
13. Neupane K, Hoffer NQ, Woodside MT (2018) Measuring the local velocity along transition paths during the folding of single biological molecules. *Phys Rev Lett* 121: 018102.
14. Pyo AGT, Hoffer NQ, Neupane K, Woodside MT (2018) Transition-path properties for folding reactions in the limit of small barriers. *J Chem Phys* 149:115101.
15. Neupane K, Hoffer NQ, Woodside MT (2018) Testing kinetic identities involving transition-path properties using single-molecule folding trajectories. *J Phys Chem B* 122:11095–11099.
16. Thirumalai D (1995) From minimal models to real proteins: Time scales for protein folding kinetics. *J Phys I* 5:1457–1467.
17. Shoemaker BA, Wang J, Wolynes PG (1999) Exploring structures in protein folding funnels with free energy functionals: The transition state ensemble. *J Mol Biol* 287: 675–694.
18. Shimada J, Shakhnovich EI (2002) The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc Natl Acad Sci USA* 99:11175–11180.
19. Chavez LL, Gosavi S, Jennings PA, Onuchic JN (2006) Multiple routes lead to the native state in the energy landscape of the β-trefoil family. *Proc Natl Acad Sci USA* 103: 10254–10258.
20. Juraszek J, Bolhuis PG (2006) Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc Natl Acad Sci USA* 103:15859–15864.
21. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA* 106:19011–19016.
22. Ensign DL, Pande VS (2009) The Fip35 WW domain folds with structural and mechanistic heterogeneity in molecular dynamics simulations. *Biophys J* 96:L53–L55.
23. Henry ER, Best RB, Eaton WA (2013) Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. *Proc Natl Acad Sci USA* 110: 17880–17885.
24. Kim WK, Netz RR (2015) The mean shape of transition and first-passage paths. *J Chem Phys* 143:224108.
25. Makarov DE (2015) Shapes of dominant transition paths from single-molecule force spectroscopy. *J Chem Phys* 143:194103.
26. Cossio P, Hummer G, Szabo A (2018) Transition paths in single-molecule force spectroscopy. *J Chem Phys* 148:123309.
27. Woodside MT, et al. (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc Natl Acad Sci USA* 103:6190–6195.
28. Woodside MT, et al. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314:1001–1004.
29. Zhang BW, Jasnow D, Zuckerman DM (2007) Transition-event durations in one-dimensional activated processes. *J Chem Phys* 126:074504.
30. Chaudhury S, Makarov DE (2010) A harmonic transition state approximation for the duration of reactive events in complex molecular rearrangements. *J Chem Phys* 133: 034118.
31. Faccioli P, Sega M, Pederiva F, Orland H (2006) Dominant pathways in protein folding. *Phys Rev Lett* 97:108101.
32. Cossio P, Hummer G, Szabo A (2015) On artifacts in single-molecule force spectroscopy. *Proc Natl Acad Sci USA* 112:14248–14253.
33. Woodside MT, Lambert J, Beach KSD (2014) Determining intrachain diffusion coefficients for biopolymer dynamics from single-molecule force spectroscopy measurements. *Biophys J* 107:1647–1653.
34. Foster DAN, et al. (2018) Probing position-dependent diffusion in folding reactions using single-molecule force spectroscopy. *Biophys J* 114:1657–1666.
35. Nam G-M, Makarov DE (2016) Extracting intrinsic dynamic parameters of biomolecular folding from single-molecule force spectroscopy experiments. *Protein Sci* 25:123–134.
36. Berezhkovskii AM, Makarov DE (2018) Single-molecule test for Markovianity of the dynamics along a reaction coordinate. *J Phys Chem Lett* 9:2190–2195.
37. Medina E, Satija R, Makarov DE (2018) Transition path times in non-Markovian activated rate processes. *J Phys Chem B* 122:11400–11413.
38. Best RB, Hummer G (2010) Coordinate-dependent diffusion in protein folding. *Proc Natl Acad Sci USA* 107:1088–1093.
39. Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. *Science* 267:1619–1620.
40. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* 4: 10–19.
41. Eaton WA, Wolynes PG (2017) Theory, simulations, and experiments show that proteins fold by multiple pathways. *Proc Natl Acad Sci USA* 114:E9759–E9760.
42. Onoa B, et al. (2003) Identifying kinetic barriers to mechanical unfolding of the *T. thermophila* ribozyme. *Science* 299:1892–1895.
43. Udgaonkar JB (2008) Multiple routes and structural heterogeneity in protein folding. *Annu Rev Biophys* 37:489–510.
44. Sen Mojumdar S, et al. (2017) Partially native intermediates mediate misfolding of SOD1 in single-molecule folding trajectories. *Nat Commun* 8:1881.
45. Roca J, et al. (2018) Monovalent ions modulate the flux through multiple folding pathways of an RNA pseudoknot. *Proc Natl Acad Sci USA* 115:E7313–E7322.
46. Radford SE, Dobson CM, Evans PA (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358:302–307.
47. Sabelko J, Ervin J, Gruebele M (1999) Observation of strange kinetics in protein folding. *Proc Natl Acad Sci USA* 96:6031–6036.
48. Guinn EJ, Jagannathan B, Marqusee S (2015) Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat Commun* 6:6861.
49. Muñoz V, Cerminara M (2016) When fast is better: Protein folding fundamentals and mechanisms from ultrafast approaches. *Biochem J* 473:2545–2559.
50. Neupane K, Woodside MT (2016) Quantifying instrumental artifacts in folding kinetics measured by single-molecule force spectroscopy. *Biophys J* 111:283–286.
51. Lu J, Nolen J (2015) Reactive trajectories and the transition path process. *Probab Theory Relat Fields* 161:195–244.
52. Berezhkovskii AM, Szabo A, Greives N, Zhou H-X (2014) Multidimensional reaction rate theory with anisotropic diffusion. *J Chem Phys* 141:204106.

PNAS
www.pnas.org

Supplementary Information for

Measuring the average shape of transition paths during the folding of a single biological molecule

Noel Q. Hoffer, Krishna Neupane, Andrew G.T. Pyo, Michael T. Woodside

Corresponding author: Michael Woodside
Email: michael.woodside@ualberta.ca

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1816602116

Table S1: Diffusion coefficients from fitting $\langle x(t|\tau_{TP})\rangle$. Errors represent standard deviation from bootstrapping analysisof fits.

Hairpin	D ($\times 10^5$ nm^2/s)		
	unfolding	refolding	average
30R50/T4	6 ± 1	6 ± 1	6 ± 1
20R100/T4	7 ± 2	6 ± 2	6 ± 2
20R55/T4	8 ± 2	8 ± 2	8 ± 2
20TS06/T4	4 ± 1	4 ± 1	4 ± 1
20R25/T4	7 ± 1	6 ± 1	6 ± 1

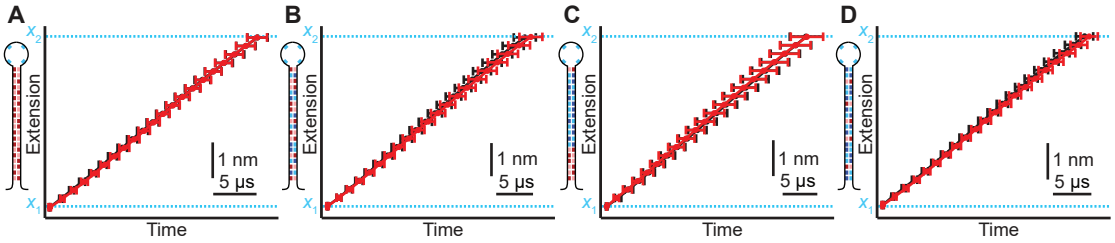

Figure S1: Time-domain average transition-path shapes $\langle t(x)\rangle$ for different hairpins. The path shape averaged in the time domain, $\langle t(x)\rangle$, is shown for unfolding (black) and refolding (red, time-reversed) of hairpins (A) 20R100/T4, (B) 20R55/T4, (C) 20TS06/T4, and (D) 20R25/T4.

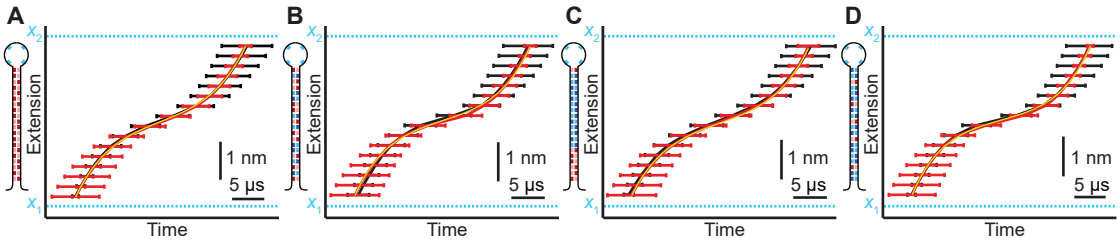

Figure S2: Extension-domain average transition-path shapes $\langle x(t|\tau)\rangle_\tau$ for different hairpins. The path shape averaged in the extension domain, $\langle x(t|\tau)\rangle_\tau$, is shown for unfolding (black) and refolding (red, time-reversed) of hairpins (A) 20R100/T4, (B) 20R55/T4, (C) 20TS06/T4, and (D) 20R25/T4. The fits to Eq. 1 (yellow) for unfolding and refolding curves are visually indistinguishable

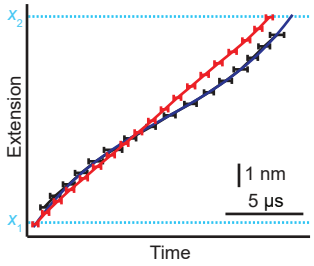

Figure S3: Effects of noise on time-domain and extension-domain averages. Computing the average path shapes $\langle t(x)\rangle$ (red) and $\langle x(t|\tau)\rangle_\tau$ (black) from an ensemble of simulated transition paths generated by adding noise sampled from a Gaussian distribution to each point on the dominant transition path defined by Eq. 1 (blue), $\langle t(x)\rangle$ became more linear, as in experimental data, whereas $\langle x(t|\tau)\rangle_\tau$ remained unaffected and matched the original analytical path shape.

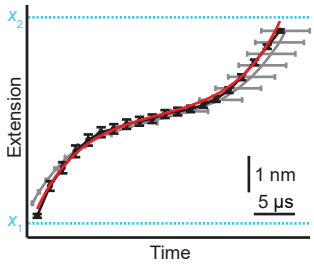

Figure S4: Extension-domain average path shape at τ_{TP}. The path shape averaged in the extension domain for transitions of the hairpin 30R50/T4 with duration equal to the average transition-path time, $\langle x(t|\tau_{TP})\rangle$ (black), is very close to the shape obtained from averaging over all transition times, $\langle x(t|\tau)\rangle_\tau$ (grey), and is well fit by Eq. 1 for the dominant path shape (red), returning a diffusion coefficient close to the result obtained from fitting $\langle x(t|\tau)\rangle_\tau$. Similar results were found for the other hairpin sequences.