



Published in final edited form as:

Science. 2006 November 10; 314(5801): 1001–1004. doi:10.1126/science.1133601.

Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid

Michael T. Woodside^{1,2}, Peter C. Anthony³, William M. Behnke-Parks², Kevan Larizadeh², Daniel Herschlag⁴, and Steven M. Block^{2,5,*}

¹National Institute for Nanotechnology, National Research Council of Canada, Edmonton AB, T6G 2M9, Canada

²Department of Biological Sciences, Stanford University, Stanford, CA 94305

³Biophysics Program, Stanford University, Stanford, CA 94305

⁴Department of Biochemistry, Stanford University, Stanford, CA 94305

⁵Department of Applied Physics, Stanford University, Stanford, CA 94305

Abstract

Nucleic acid hairpins provide a powerful model system for understanding macromolecular folding, with free energy landscapes that can be readily manipulated by changing the hairpin sequence. The full shapes of energy landscapes for the reversible folding of DNA hairpins under controlled loads exerted by an optical force clamp were obtained by deconvolution from high-resolution, single-molecule trajectories. The locations and heights of the energy barriers for hairpin folding could be tuned by adjusting the number and location of G:C base pairs, and the presence and position of folding intermediates were controlled by introducing single-nucleotide mismatches.

Elucidating the mechanisms by which proteins and nucleic acids fold into three-dimensional structures is key to developing insights into biomolecular function (1), improving predictive models (2,3), and understanding the basis of diseases linked to misfolding (4). For over two decades, free energy landscape formalisms have provided the fundamental conceptual framework for describing folding (5). Numerous experimental and theoretical studies have probed specific features of folding landscapes, including the properties of transition states (6), intermediate states (7), and the ruggedness of the energy surface (8). Experiments, however, have characterized only limited aspects of the folding landscape, such as the locations and heights of energy barriers, and how these barriers change when perturbed by solvent substitutions, temperature jumps, substrate changes, or mutations (9). Direct measurements of the shape of an energy landscape at all points along the reaction coordinate have not been feasible. Here, we show how the full energy landscape for the formation of a nucleic acid hairpin can be derived from sufficiently high-resolution trajectories of single-molecule folding.

Single nucleic acid hairpins subjected to mechanical loads provide a powerful model system for investigating energy landscapes and understanding the effects of primary and secondary

*To whom correspondence should be addressed. Email: sblock@stanford.edu.

Supporting Online Material

www.sciencemag.org

Materials and Methods

Figs. S1, S2

Tables S1, S2

References

structure on folding (10-13). The molecular end-to-end extension is recorded during the folding transition, supplying a natural reaction coordinate that can be related directly to the number of bases paired in the hairpin stem. Previous work has characterized specific aspects of the folding landscape. In particular, short hairpins tend to fold as simple, two-state systems (10,13), indicative of a single transition energy barrier. Conventional analysis of single-molecule records supplies the free energy difference between the folded and unfolded states, as well as the height and location of the barrier (10). For hairpins with random (unpatterned) stem sequences, the barrier is typically located close to the unfolded state, with a height controlled largely by the size of the loop (13). However, finer details of the folding landscape have been heretofore inaccessible, due to limited spatial and temporal resolution as well as instrumental baseline drift. Using a high-bandwidth, passive force clamp with an ultra-stable dumbbell assay (14), we have now been able to reconstruct the shape of the landscape.

Sets of DNA hairpins were synthesized in which the heights and locations of energy barriers were systematically varied, as well as the numbers and locations of any folding intermediates. Sequences were designed based on a model of the sequence-dependent energy landscape derived from the thermodynamic and mechanical properties of nucleic acids (13,15). Both ends of the hairpins were attached to long handles of dsDNA (13) bound specifically to polystyrene beads held in a dumbbell configuration by two independently-controlled optical traps (Fig. 1A). A constant force, F , was applied with a force clamp (14), and high-resolution trajectories of the end-to-end extension ($\sim 0.1 \text{ nm}/\sqrt{\text{Hz}}$) were recorded for a range of forces. The extensions of folded, unfolded, and any intermediate states were measured directly from these records. The locations and heights of energy barriers between these states were computed from the force-dependence of the state lifetimes (10,13). These measurements of specific points on the landscape were then taken as benchmarks for an experimental determination of the free energy at every point along the reaction coordinate, deconvolving the measured probability distribution of hairpin extension to correct for blurring effects arising from thermal motions associated with the beads and the DNA tether.

A typical record of extension under load (Fig. 1B) shows two-state folding behavior: two nearly Gaussian peaks in the extension histogram correspond to the folded and unfolded states. Here, $F \approx F_{1/2}$, the load at which the hairpin spends equal time in each state. The lifetimes of the folded (τ_f) and unfolded (τ_u) states depend exponentially on F according to $\tau_f(F) \propto \exp(-F\Delta x_f^\ddagger/k_B T)$, where Δx_f^\ddagger is the distance to the barrier from the folded state and $k_B T$ is the thermal energy (16); an analogous expression holds for $\tau_u(F)$. Previously, the transition state (TS) for folding a hairpin with an unpatterned stem sequence was found to involve the formation of 1–2 base pairs adjacent to the loop (13), resulting in an energy barrier near the unfolded state. In contrast, the barrier for the hairpin in Fig. 1B lies much closer to the folded ($\Delta x_f^\ddagger = 5.4 \pm 0.5 \text{ nm}$) than to the unfolded state ($\Delta x_u^\ddagger = 13 \pm 1 \text{ nm}$), implying that the TS requires the formation of ~ 15 base pairs. This difference in behavior is due to the particular sequence selected for the hairpin: a contiguous block of strong G:C base pairs placed near the base of an A:T-rich stem moves the barrier near to the folded state (Fig. 1B, inset).

We created a family of hairpins in which the TS position was systematically manipulated by moving the block of G:C base pairs to various locations within the stem (Table S1). Determining the barrier location for each hairpin, we found that the TS moved in concert with the G:C block, always located at the edge of the block nearest the loop (Fig. 1C; Table S2). The sum of Δx_f^\ddagger and Δx_u^\ddagger agreed well with the distance between folded and unfolded states (Δx), as expected for a pure two-state system. Measurements were in excellent agreement with landscape model predictions (Fig. 1C).

We also created a family of hairpins where the barrier position was fixed at the center of the stem using a G:C block, but the barrier height was altered by changing the overall stem G:C

content (Table S1). State lifetimes measured at $F_{1/2}$ for each hairpin (17) varied by a factor of ~ 600 over the entire family (Fig. 1D; Table S2). Assuming Arrhenius behavior for the lifetimes, $\tau_{1/2} = \tau_0 \exp(\Delta G_{1/2}^\ddagger/k_B T)$, where $\Delta G_{1/2}^\ddagger$ is the barrier height at $F_{1/2}$ and τ_0^{-1} the attempt rate at zero force, these results show that the barrier height changed by $6.4 \pm 0.7 k_B T$, matching the variation predicted by the model, $6.9 \pm 0.2 k_B T$. All together, the results of Fig. 1 confirm the remarkable level of control afforded by this system over the folding landscape: the TS can be placed at will along the reaction coordinate and its energy adjusted over a wide range simply by manipulating the hairpin sequence.

Extension records such as those in Fig. 1B and (10,12,13) have traditionally been interpreted in terms of two-state folding over a single energy barrier. However, not all hairpin sequences exhibit strict, two-state folding behavior. For example, it was recently reported that the nominal “folded state” may, in fact, consist of an ensemble of states comprised of the folded state plus a series of frayed states with one or more base pairs unzipped (13). Moreover, short-lived intermediate states may be present that are unobservable at the available temporal resolution (18). Simple two-state analysis ignores details of the trajectory between folded and unfolded states because the motion is taken to be instantaneous, and properties of the energy landscape are inferred only from characteristics of the two states. To induce the hairpin to spend more time between folded and unfolded states, and to observe intermediate properties during folding more clearly, we manipulated the sequence to produce a local potential well between the folded and unfolded states by inserting a single T:T mismatch at various positions along the hairpin stem.

When the mismatch was placed at the seventh base pair from the base of the stem, extension records at $F \approx F_{1/2}$ revealed a shoulder on the histogram peak at low extension (nominally, the folded state), indicative of a third peak representing an intermediate state (Fig. 2A). Such a shoulder was observed systematically in all records, but associated only with the low-extension peak. The existence of an intermediate state may also be inferred from the rapid fluctuation of ~ 5 nm amplitude recorded at low extensions (Fig. 2B). Similar results were obtained by introducing a G:A or a G:T mismatch, rather than T:T. Repeating these measurements for an entire family of hairpins with mismatches located 4–16 bp from the base of the stem (Table S1), we consistently observed the signature of an intermediate state, whose distance from the folded and unfolded states depended on the location of the mismatch (Fig. 2C; Table S2). Interpreting the hairpin extension in terms of the number of base pairs formed, the intermediate states correspond to hairpins partially folded up to the point of mismatch.

These results demonstrate the precision with which certain features of the folding landscape can be determined, but they define only a few key points on the energy landscape. Notably, they don't address more general features of the landscape, such as the widths and curvatures of the potential wells or barriers, which are known to affect folding (19). By further analyzing the folding trajectories, however, the entire landscape along the reaction coordinate can be reconstituted. The free energy at a given extension, $\Delta G(x)$, is related to the probability density, $P(x)$, through $\Delta G(x) = -k_B T \ln[P(x)]$ (20). Although conceptually straightforward, this method of determining $\Delta G(x)$ requires accurate measurements of $P(x)$ in the region between the states, where the hairpin spends little time (~ 100 – 300 μ s, here). Hundreds to thousands of transitions must therefore be sampled at high bandwidth, necessitating exceptional instrumental stability. A second complication is that the measured extension represents that of a hairpin attached to dsDNA handles and beads, rather than an isolated hairpin. The thermal and mechanical properties of the trapped dumbbell smooth and dampen the apparent motions of the hairpin (13). The underlying energy landscape may be recovered from $P(x)$, however, by a deconvolution process.

To reconstruct the full energy landscape, we measured the folding trajectory of single hairpins at $F \approx F_{1/2}$ at high bandwidth (50 kHz) for 5–15 min and created a histogram of the extension, $P(x)$. Instrumental drift was typically ≤ 1 nm. The point spread function (PSF) for the deconvolution, $S(x)$, was estimated from extension histograms of the folded state for a hairpin with 100% stem G:C content and found to be a Gaussian (Fig. S2) whose width is governed by the stiffness of the trapped dumbbell (14). The energy landscape was then determined by a constrained nonlinear iterative deconvolution (21) of the extension histogram. An initial guess for the potential, $\Delta G^{(0)}(x)$, was constructed by assuming parabolic potential wells located at the histogram maxima, separated by a parabolic barrier whose height and position were determined from the measured, force-dependent rates (as in Fig. 1). The associated extension probability $p^{(0)}(x)$ was then convolved with the PSF and compared to $P(x)$. The difference was subtracted from $p^{(0)}(x)$, constraining the probability to be between 0 and 1, and the process was iterated (15). The solutions, $p^{(n)}(x)$, and associated landscapes, $\Delta G^{(n)}(x)$, are shown in Fig. 3, along with the measured $P(x)$, $\Delta G(x)$, and residuals $R(x) = P(x) - S(x) \otimes p^{(n)}(x)$, for four different hairpin sequences designed to explore a range of barrier positions, heights, and possible intermediate states. Shown in Fig. 3 are a 20 bp stem with a TS located 18 bp from the base of the stem (Fig. 3A), a 20 bp stem with TS located 6 bp from the base of the stem (Fig. 3B), a 20 bp stem with T:T mismatch located 7 bp from the end of the stem (Fig. 3C), and an unpatterned stem sequence of length 30 bp (Fig. 3D). In all four cases, the deconvolution algorithm generated a stable solution with acceptably small residuals.

The subtle differences seen in $P(x)$ and $\Delta G(x)$ were sharpened by the deconvolution procedure into recognizably different landscapes that reflected the underlying sequence and recapitulated the results in Figs. 1 and 2. In Fig. 3A, the barrier is located near the unfolded state, whereas in Fig. 3B it is nearer the folded state. The hairpin in Fig. 3C, which contains a mismatch, shows a clearly-resolved intermediate state, corresponding to partial folding up to the point of mismatch. These measurements go beyond the previous results, however, by revealing details of the well and barrier shapes. For example, the energy minimum for the folded state in Fig. 3A is significantly broader than that in Fig. 3B: the width of this well supplies direct evidence that the nominally folded state for this hairpin actually consists of an ensemble of states with up to ~ 4 bp unzipped. A similar situation is seen in Fig. 3D, although the slightly steeper well suggests that the fully folded state plays a more dominant role in this mixed ensemble than it does in Fig. 3A. The energy barriers in Figs. 3A and D are clearly different: the barrier in Fig. 3D is wider than in Fig. 3A, indicating a TS that is less well-defined and therefore more susceptible to experimental perturbation, *e.g.*, by mutagenesis or solvent condition changes.

To explore the validity of these measurements, we compared the experimental landscapes with predictions of the model (Fig. 3). We found excellent agreement across the entire landscape for all hairpins studied (within $< 1 k_B T$), with two exceptions. At the lowest extensions, corresponding to regions deconvolved from physical compressions of the double helix (which can arise from thermal fluctuations) as well as elongations, the experimental potential is systematically less stiff than the model. This discrepancy may be attributable to an inaccurate description of the confining potential, somewhat arbitrarily taken to be a Morse potential (22). In addition, the barrier for exiting the unfolded state in Fig. 3B rises to the TS more slowly than predicted, lagging by up to $3 k_B T$ at the point of greatest discrepancy. We speculate that this deviation may result from the large number of base pairs that must be formed to reach the TS from the unfolded state, which allows more opportunities for abortive refolding attempts involving misfolded base pairs. In principle, the sequence for this particular hairpin allows for a number of misfolded states containing short, 2–3 bp helices. Any such misfolding, neglected in the model, would tend to increase the probability of extensions near the unfolded state, exactly as observed.

The deconvolution approach described here has known limitations. To obtain adequate statistics, folding must occur sufficiently frequently that large numbers of transitions can be recorded. In the present case, this places a practical limit on the folding rate of $\sim 0.1 \text{ s}^{-1}$, which is faster than some slow folding transitions found in proteins or ribozymes. The numerical stability of any deconvolution process depends on the quality of the input data (both for the record being analyzed and the PSF employed). In practice, only a limited range of frequency information can be recovered by deconvolution, restricting the resolution of the reconstructed landscape, particularly at the shortest length scales (23). Moreover, experimental noise may become amplified by deconvolution, producing artifactual features that further complicate determinations of short-scale behavior (21). The challenges posed by deconvolution, however, may be mitigated by increasing the stiffness of the experimental system, which reduces the smoothing of trajectories (24). Improvements may be achieved by increasing the stiffness of the handles (*e.g.*, by making them shorter or from materials other than dsDNA). Application of the approach described here to peptides or more complex nucleic acid sequences may supply further insights into how energy landscapes guide the folding process.

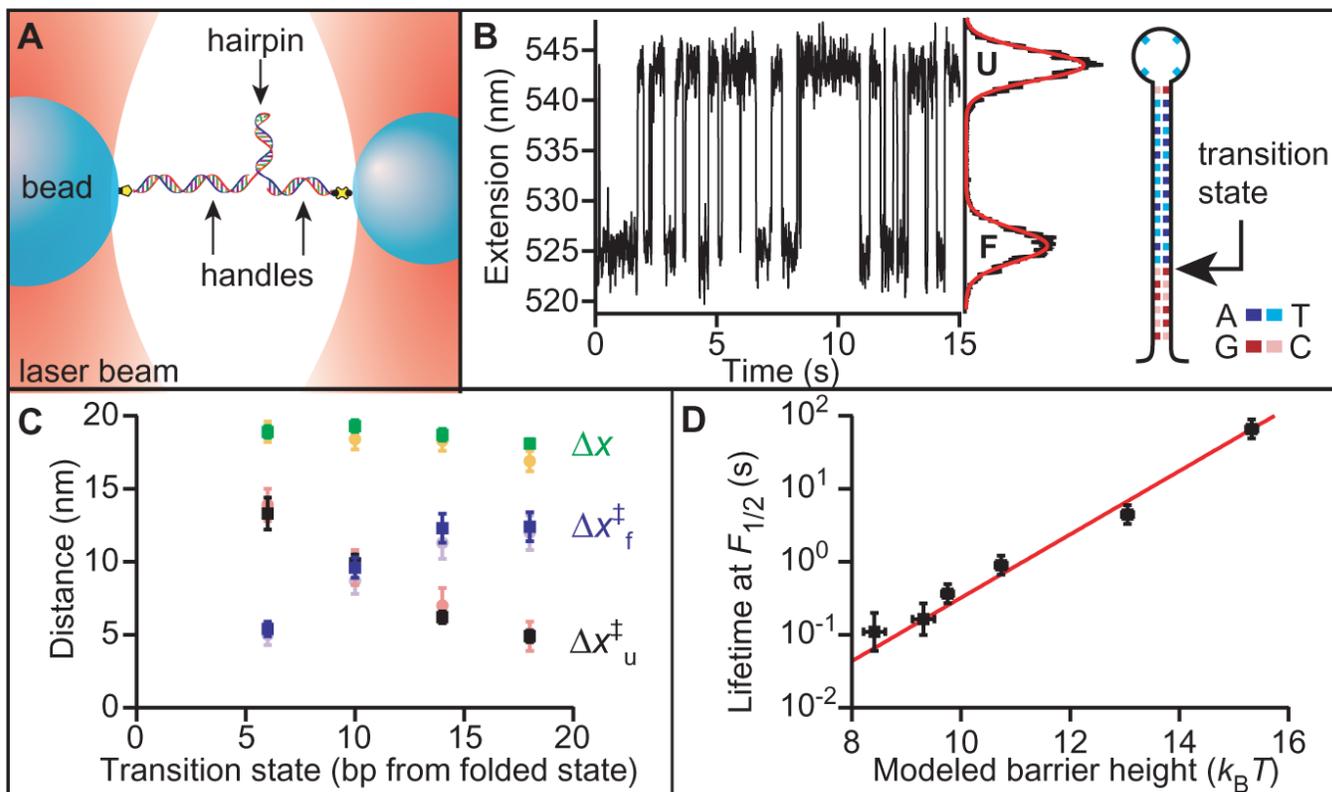
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

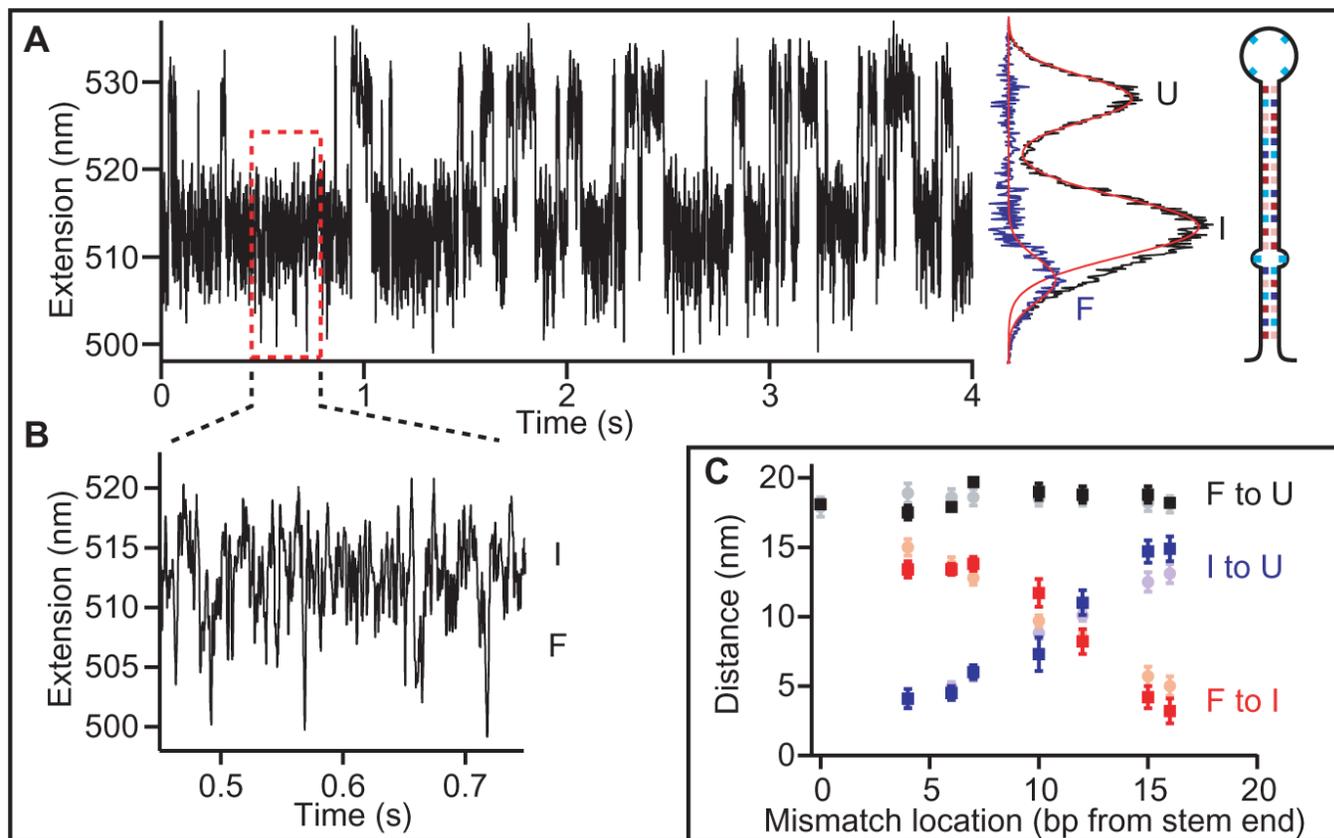
References and Notes

1. Fersht, A. Structure and Mechanism in Protein Science. Freeman, WH., editor. New York: 1999.
2. Petrey D, Honig B. Mol. Cell 2005;20:811. [PubMed: 16364908]
3. Bradley P, Misura KMS, Baker D. Science 2005;309:1868. [PubMed: 16166519]
4. Dobson CM. Nature 2003;426:884. [PubMed: 14685248]
5. Onuchic JN, Wolynes PG. Curr. Opin. Struct. Biol 2004;14:70. [PubMed: 15102452]
6. Daggett V, Fersht AR. Trends in Biochem. Sci 2003;28:18. [PubMed: 12517448]
7. Kim PS, Baldwin RL. Annu. Rev. Biochem 1990;59:631. [PubMed: 2197986]
8. Hyeon C, Thirumalai D. Proc. Natl. Acad. Sci. USA 2003;100:10249. [PubMed: 12934020]
9. Buchner, J.; Kiefhaber, T., editors. Protein Folding Handbook. Wiley-VCH; Weinheim: 2005.
10. Liphardt J, Onoa B, Smith SB, Tinoco I Jr. Bustamante C. Science 2001;292:733. [PubMed: 11326101]
11. Onoa B, et al. Science 2003;299:1892. [PubMed: 12649482]
12. Li PTX, Collin D, Smith SB, Bustamante C, Tinoco I Jr. Biophys. J 2006;90:250. [PubMed: 16214869]
13. Woodside MT, et al. Proc. Natl. Acad. Sci. USA 2006;103:6190. [PubMed: 16606839]
14. Greenleaf WJ, Woodside MT, Abbondanzieri EA, Block SM. Phys. Rev. Lett 2005;95:208102. [PubMed: 16384102]
15. Materials and Methods are available as supporting material on *Science* Online.
16. Bell GI. Science 1978;200:618. [PubMed: 347575]
17. Because all lifetimes were measured at $F_{1/2}$, the folded and unfolded states were nearly isoenergetic for every hairpin, and Hammond behavior was neither expected nor observed
18. Fernandez JM, Chu S, Oberhauser AF. Science 2001;292:653. [PubMed: 11330326]
19. Kramers HA. Physica 1940;7:284.
20. McCann LI, Dykman M, Golding B. Nature 1999;402:785. Experimentally-determined probability densities have been used to measure the energy landscape for a μm -scale bead hopping between adjacent optical traps. Jeney S, Stelzer EHK, Grubmüller H, Florin E-F. ChemPhysChem 2005;5:1150. [PubMed: 15446737] to probe the elasticity of kinesin molecules. Walther KA, Brujić J, Li H, Fernandez JM. Biophys. J 2006;90:3806. [PubMed: 16500987] to support the existence of otherwise hidden intermediate states based on thermal fluctuations but not to measure the energy landscape for a molecular transition

21. Jansson, PA., editor. Deconvolution of Images and Spectra. Vol. 2nd ed. Academic Press; New York: 1997.
22. A likely explanation for the weak confining potential observed experimentally is that the hairpin can rotate, effectively producing extensions lower than the folded state as the handles move past one another, driven by thermal fluctuations
23. In practice, we found a resolution limit of ~ 2 nm, which includes the effects of instrumental drift
24. The data in Figs. 3A-C were measured in an unclamped configuration for this reason, because passive force-clamping reduces the system stiffness (14)
25. We thank members of the Block lab and the Ribozyme Folding Project for helpful discussions and manuscript comments. Supported by NIH P01-GM066275

**Fig. 1.**

(A) Cartoon of the experimental geometry, showing a DNA hairpin attached at each end to dsDNA handles bound to two optically trapped beads (not to scale). (B) Extension record of a hairpin designed with an energy barrier 6 bp from the base of the stem (sequence in inset) displays two states, folded (F) and unfolded (U). (C) Unfolding distance Δx (green) and distance to barrier from unfolded state Δx_u^\ddagger (black) and folded state Δx_f^\ddagger (blue), as a function of the expected transition state location. Data (shown with std. errs.) agree well with model predictions for Δx (yellow), Δx_u^\ddagger (red), and Δx_f^\ddagger (purple). (D) Lifetime of folded/unfolded states at $F_{1/2}$ as a function of the expected barrier height. The lifetime rises exponentially with the modeled barrier height, with slope $(k_B T)^{-1}$ (red line).

**Fig. 2.**

(A) Extension record of hairpin with a T:T mismatch 7 bp from the base of the stem (sequence in inset). Fit of extension histogram to two Gaussian peaks reveals a third Gaussian residual (fits in red, residual in blue), indicating three states: unfolded (U), intermediate (I), and folded (F). (B) Blow-up of the extension record shows rapid, ~5-nm fluctuations between F and I states. (C) Distance from F to U (black), F to I (red), and I to U (blue) plotted versus the mismatch location. The intermediate state location moves in concert with the mismatch, in good agreement with model predictions (light-colored solid circles in grey, orange, purple), indicating that the intermediate state consists of a hairpin folded up to the mismatch.

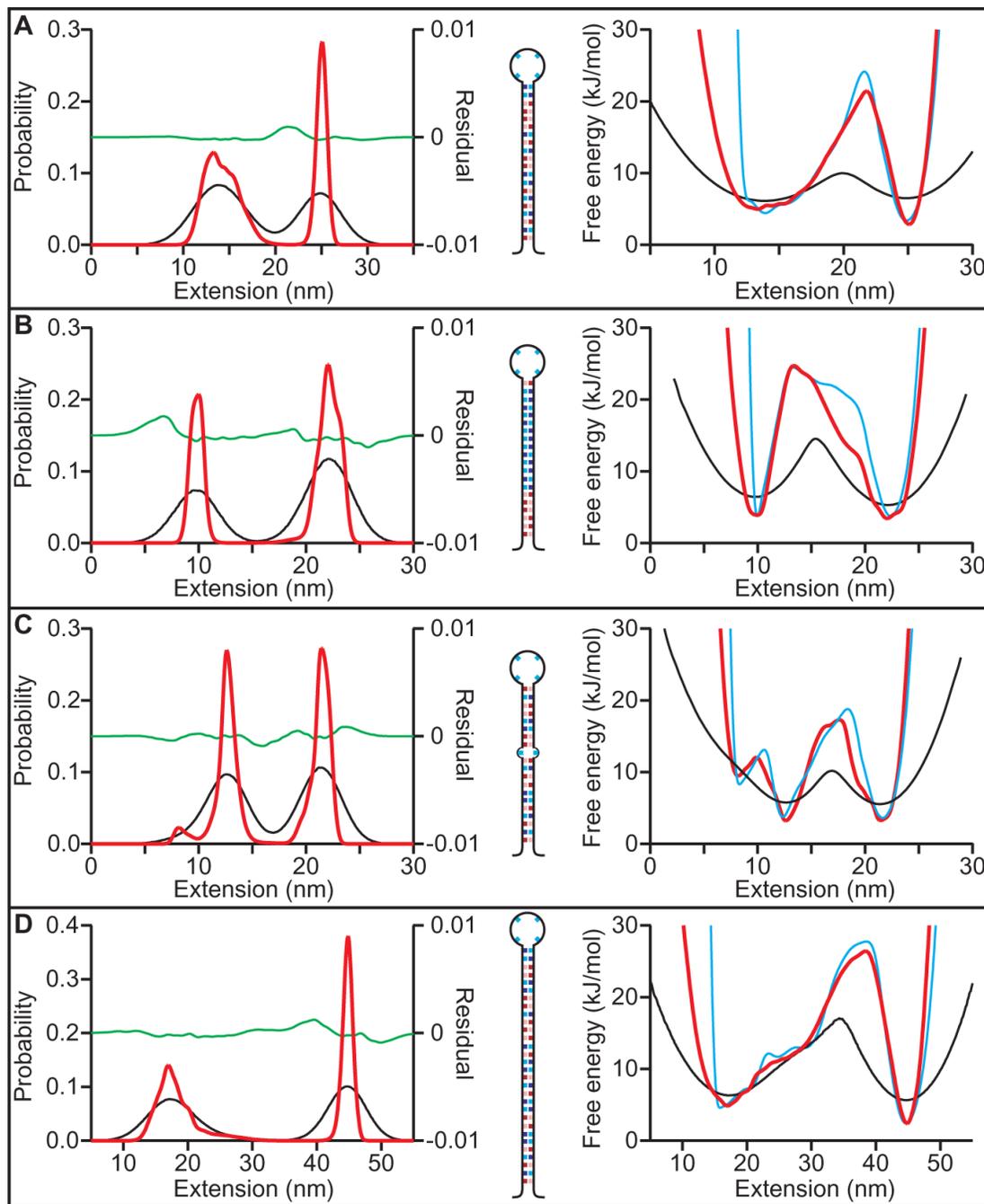


Fig. 3. Potential landscapes determined by deconvolution of the extension histograms based on thousands of folding transitions. Experimental probability distributions and associated free energy landscapes (black) were deconvolved to remove the effects of blurring caused by thermal motions of beads attached to the hairpin via elastic handles. Probability distributions (left column) and landscapes (right column) recovered by Gaussian deconvolution (red) have small residual errors (green), and agree well with energy landscapes predicted by the model (blue). **(A)** Hairpin with a transition state 18 bp from the base of the stem. **(B)** Hairpin with a transition state 6 bp from the base of the stem. **(C)** Hairpin with a T:T mismatch 7 bp from the base of the stem. **(D)** Hairpin with an unpatterned, 30-bp stem sequence.