www.sciencemag.org/cgi/content/full/314/5801/1001/DC1

Supporting Online Material for

Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid

Michael T. Woodside, Peter C. Anthony, William M. Behnke-Parks, Kevan Larizadeh, Daniel Herschlag, Steven M. Block*

*To whom correspondence should be addressed. E-mail: sblock@stanford.edu

Published 10 November 2006, *Science* **314**, 1001 (2006).
DOI: 10.1126/science.1133601

This PDF file includes

Materials and Methods

Sample preparation. Hairpin constructs were made as described previously (*S1*). Briefly, these each consisted of a hairpin sequence connected to a 621-bp, digoxigenin-labeled dsDNA handle on the 3′ end, and a 1036-bp, biotin-labeled dsDNA handle on the 5′ end. The hairpin sequence was separated from each handle by an abasic site (a deoxyribose spacer), inserted to reduce stacking interactions between the handles and the hairpin. These constructs were incubated with 600-nm diameter polystyrene beads labeled with avidin and 730-nm diameter beads labeled with anti-digoxigenin to produce "dumbbells," as shown in Fig. 1A (*S2*). Dumbbells were diluted in assay buffer (50 mM MOPS pH 7.5, 200 mM KCl) with 2% oxygen scavenging system (250 mg/mL glucose, 37 mg/mL glucose oxidase, 1.7 mg/mL catalase), and introduced into a flow cell.

Measurement and analysis. All measurements were made at 23 ± 0.5°C in an optical trap with two trapping beams and two detector beams, as described previously (*S3*). The positions and intensities of the two orthogonally-polarized infrared trap beams were controlled independently by acousto-optic deflectors, while the positions of the two beads were detected independently by light from two orthogonally-polarized red detector beams that was scattered by the beads onto position sensitive diodes. The stiffnesses of the traps were calibrated using standard techniques, and the position of the bead in each trap was calibrated for every dumbbell (*S4*). Where not otherwise indicated, measurements were made using a passive force clamp, described elsewhere (*S3*), which maintains constant force during the motions of the hairpin. Records were measured at constant force for 10–3000 s, depending on the hairpin folding rate. Data were sampled at bandwidths ranging from 1–50 kHz and Bessel-filtered online at the appropriate Nyquist frequency. Data used for deconvolution of the energy landscape were not filtered offline, while the rest of the data were median-filtered with a 1–200 ms window. Folded and unfolded states for hairpins with two-state behavior were partitioned automatically using a software threshold adjusted to the extension midway between the two states.

Hairpin sequences. All hairpins (except one) had a 20-bp stem, and all had a 4-nt poly-thymidine loop (a tetraloop sequence chosen to minimize any intraloop stacking and structure). The sequences of the hairpins are listed in Table S1 below. The hairpins in the family with varying transition state locations were named "20TSxx/T4," where "20" indicates the length of the stem, "TSxx" indicates a transition state located putatively "xx" bp from the end of the stem (predicted by the landscape model used in the design process), and "T4" indicates the sequence of the loop. The hairpins with varying barrier heights were named "20Hyy/T4", where "Hyy" indicates a barrier height predicted by the landscape model to be "yy" kJ/mol. The hairpins with varying mismatch locations were named "20Mzz_nn/T4", where "Mzz" indicates a mismatch located "zz" base pairs from the base of the stem and "nn" indicates the mismatch base identity. Two hairpins had sequences described previously: the hairpin in the 'mismatch' family that had no mismatch was hairpin 20R55/T4 from ref. S1 (where "R55" indicates a random stem sequence with 55% G:C content), while the hairpin with a 30-bp stem shown in Fig. 3A was hairpin 30R50/T4 from ref. S1.

Effects of force on folding landscape. The application of force constrains the unfolded state and partially unfolded intermediates to adopt a low-entropy, extended conformation, justifying the use of a one-dimensional description of the landscape. The externally-applied force tilts the energy landscape because of the work performed by the optical trap on the system. The precise effect of force on the landscape depends on how the force changes during the folding reaction itself, as described elsewhere (*S3*). For constant force, as in most measurements presented here,

the situation is described by Fig. S1: the energy of the unfolded state decreases by $F\Delta x$ while the energy of the transition state decreases by $F\Delta x^{\ddagger}$. We assume that the position of the transition state remains invariant over the small range of force applied here, an approximation valid in the limit that the curvature of the potential barrier is much greater than the change in the force.

Landscape model. The model used to predict the hairpin folding behavior has been described in detail previously (*S1*). Briefly, this model calculates the free energy required to unzip each base pair using nearest neighbor energies from the program *mfold* 3.1 (*S5*), and adds to this the free energy required to stretch the unzipped ssDNA using a worm-like chain approximation (*S6*) and the work done by the trap, yielding the free energy as a function of the extension of the hairpin and the force applied by the trap, as follows:

$$\Delta G(F,x) = \Delta G_{\text{unzip}}(x) + \Delta G_{\text{stretch}}(x) + \int_0^x F(\xi)\,d\xi \qquad , x \geq 0;$$

$$\varepsilon\left[1-\exp\left(-x/a\right)\right]^2 + \int_0^x F\left(\xi\right)d\xi \qquad , x < 0, \qquad (1)$$

$$\text{where } \Delta G_{\text{stretch}}\left(x\right) = \frac{k_{\text{B}}T}{L_{\text{p}}}\frac{L_0}{4(1-x/L_0)}\left[3(x/L_0)^2 - 2(x/L_0)^3\right].$$

A repulsive Morse potential, where ε represents the potential depth and *a* represent the well width, is used to approximate the resistance of the dsDNA helix to compression (*S7*), which can occur through thermal fluctuations. L_{p} is the persistence length and L_0 is the contour length of the ssDNA in the unfolded hairpin. This free-energy function is then smoothed by an amount corresponding to the stiffness of the unzipped ssDNA, to incorporate the effect of ssDNA elasticity, resulting in model landscapes of the form shown in Fig. 3. The heights and locations of any energy barriers were determined directly from the modeled landscape. Extension histograms were simulated by calculating the probability distribution function for the extension from the free energy landscape and convolving this with a Gaussian of appropriate width based on the elastic stiffness of the dsDNA handles. Distances between folded and unfolded states were determined from these modeled histograms. The locations of any intermediate states were determined from the positions of local minima in the energy landscape.

We note that this model contains no free fitting parameters. The model employs, as fixed parameters, values for the persistence length and the contour length of ssDNA, and for the extension of ssDNA at given force, that fall in a range consistent with previous measurements.

Numerical data. Table S2 displays the numerical values of the experimental data and model results displayed in Figs. 1 and 2. Error bars on experimental data indicate the standard error on the mean added in quadrature to the estimated systematic error arising from calibration uncertainties and the known dispersion in bead sizes. Uncertainties in the model reflect the standard deviation of the results obtained from calculations using a range of input parameters consistent with previous work (see above).

Deconvolution procedure. To reconstruct the free energy landscape, we removed by deconvolution the effects of the thermal motion of the beads attached by elastic dsDNA handles to the hairpin (which blur the extension histogram), following the non-linear constrained iterative methods described by Jansson (*S8*). Given a point-spread function $S(x)$ smoothing the true extension probability function to produce the measured extension probability $P(x)$, we construct an initial solution $p^{(0)}(x)$ and approach the true distribution function iteratively:

$$p^{(k+1)}(x) = p^{(k)}(x) + r\left[p^{(k)}(x)\right]\left(P(x) - S(x) \otimes p^{(k)}(x)\right); \quad (2)$$
$$r\left[p^{(k)}(x)\right] = r_0\left(1 - 2\left|p^{(k)}(x) - \tfrac{1}{2}\right|\right),$$

where k refers to the index of the iteration. The relaxation function $r\,[p^{(k)}(x)]$ constrains the solution to remain within the physical boundaries $0 \le p^{(k)}(x) \le 1$, with the amplitude r_0 controlling the speed of convergence. We used $r_0 = 2$ with ~300 iterations. To reduce artifactual fluctuations in the deconvolution, $P(x)$ was smoothed in a 1-nm window, as was the final solution $p^{(n)}(x)$.

$S(x)$ was estimated from histograms of the extension of a hairpin with 100% stem G:C content studied previously (hairpin 20R100/T4 from ref. S1). This hairpin is known to remain in the folded state for the forces used in the experiment of Fig. 3 (~11-14 pN). $S(x)$ was determined to be a Gaussian with a characteristic width that depended on the stiffness of the trapped dumbbell (Fig. S2). To increase the resolution of the deconvolved landscape in Fig. 3A-C, the stiffness of the trapped dumbbell was increased by measuring without a force clamp (*S3*). In these measurements, therefore, the force on the hairpins changed somewhat as the hairpins folded and unfolded. This effect modifies the energy landscape by changing the integral for the mechanical work carried out by the trap in Eq. S1. A correction to the energy landscape was therefore calculated from the known local stiffness of the trap and applied, as described previously (*S3*).

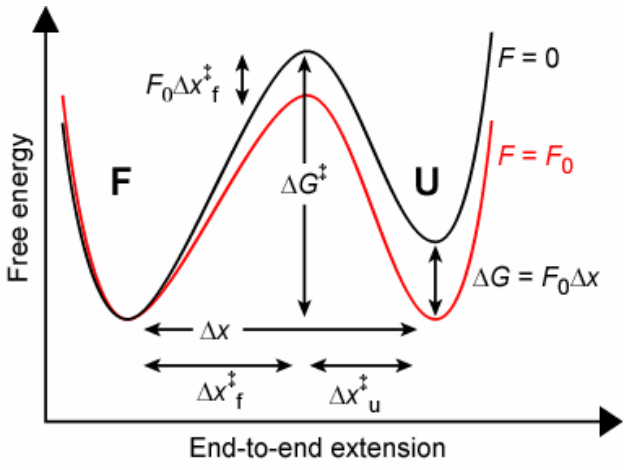

Fig. S1. Effect of force on the energy landscape. A force F_0 that is constant during the folding transition tilts the landscape uniformly, reducing the energy of the unfolded state (U) by $F_0\Delta x$, and reducing the energy of the transition state by $F_0\Delta x^{\ddagger}_f$, where Δx is the distance between F (folded state) and U, and $\Delta x^{\ddagger}_f$ is the distance between F and the energy barrier.

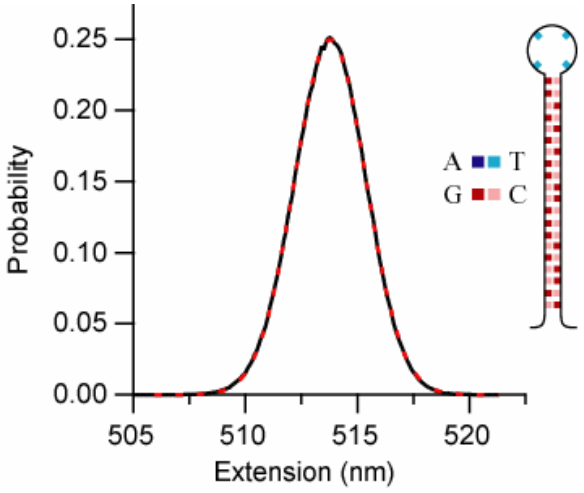

Fig. S2. A histogram of the extension of a hairpin with 100% G:C content in the stem (sequence shown in inset), which remains fully folded under 11 pN load (solid black line). The data are well fit by a Gaussian (dashed red line), experimentally confirming the Gaussian form of the point spread function (PSF) used for the deconvolution procedure.

Table S1. Sequences of the hairpins measured. The hairpin nomenclature is described in the Materials and Methods. *N* is the number of hairpin molecules measured.

Hairpin name	**Sequence**	***N***
Varying transition state location (indicated in bold italic)		
20TS06/T4	GCCGG***C***TATTATTTATATTC(T)$_4$GAATATAAATAATA***G***CCGGC	8
20TS10/T4	GACTGAAGC***G***TATTATATTA(T)$_4$TAATATAATA***C***GCTTCAGTC	4
20TS14/T4	GACTGTTCATCGG***C***TATTTA(T)$_4$TAAATA***G***CCGATGAACAGTC	7
20TS18/T4	GACTGTGTAGGTCACGG***C***TA(T)$_4$TA***G***CCGTGACCTACACAGTC	4
Varying barrier height , transition state at 10th bp in stem		
20H21/T4	GATTGTAGGCTATTATATTA(T)$_4$TAATATAATAGCCTACAATC	4
20H23/T4	CTGTCAGAGCTATTATATTA(T)$_4$TAATATAATAGCTCTGACAG	6
20H26/T4	CTGTCACGGCTATTATATTA(T)$_4$TAATATAATAGCCGTGACAG	2
20H32/T4	GCGCAACGGCTATTATATTA(T)$_4$TAATATAATAGCCGTTGCGC	5
20H38/T4	GCGCGGGCGCTTTTATATAT(T)$_4$ATATATAAAAGCGCCCGCGC	4
Varying mismatch location (indicated in bold italic)		
20R55/T4	GAGTCAACGTCTGGATCCTG(T)$_4$CAGGATCCAGACGTTGACTC	10
20M04_TT/T4	GAG***T***CAACGTCTGGATCCTG(T)$_4$CAGGATCCAGACGTTG***T***CTC	3
20M06_TT/T4	GAGTC***T***ACGTCTGGATCCTG(T)$_4$CAGGATCCAGACGT***T***GACTC	6
20M07_TT/T4	GAGTCA***T***CGTCTGGATCCTG(T)$_4$CAGGATCCAGACG***T***TGACTC	6
20M10_TT/T4	GAGTCAACG***T***CTGGATCCTG(T)$_4$CAGGATCCAG***T***CGTTGACTC	2
20M12_TT/T4	GAGTCAACGTC***T***GGATCCTG(T)$_4$CAGGATCC***T***GACGTTGACTC	6
20M15_TT/T4	GAGTCAACGTCTGG***T***TCCTG(T)$_4$CAGGA***T***CCAGACGTTGACTC	4
20M16_TT/T4	GAGTCTACGTCTGGA***T***CCTG(T)$_4$CAGG***T***TCCAGACGTAGACTC	2
20M06_AA/T4	GAGTC***A***ACGTCTGGATCCTA(T)$_4$TAGGATCCAGACGT***A***GACTC	3
20M06_GT/T4	GAGTC***G***ACGTCTGGATCCTA(T)$_4$TAGGATCCAGACGT***T***GACTC	3
Deconvolution		
30R50/T4	GAGTCAACGTACTGATCACGCTGGATCCTA(T)$_4$TAGGATCCAGCG TGATCAGTACGTTGACTC	6

Table S2. Numerical results from experiment and model. Experimental data are listed above, with model results shown directly below in italics.

Hairpins with varying transition state location (Fig. 1C)			
Hairpin name	**Δx (nm)**	**$\Delta x_u^{\ddagger}$ (nm)**	**$\Delta x_f^{\ddagger}$ (nm)**
20TS06/T4	18.9 ± 0.4	13.3 ± 1	5.4 ± 0.5
	18.9 ± 0.7	*13.9 ± 1.1*	*5.0 ± 0.7*
20TS10/T4	19.3 ± 0.4	9.9 ± 0.6	9.6 ± 0.7
	18.4 ± 0.7	*9.6 ± 1.2*	*8.7 ± 0.9*
20TS14/T4	18.7 ± 0.4	6.2 ± 0.4	12.3 ± 1
	18.3 ± 0.7	*7.0 ± 1.2*	*11.3 ± 1.1*
20TS18/T4	18.1 ± 0.3	4.9 ± 0.4	12.4 ± 1
	16.9 ± 0.7	*4.9 ± 1.0*	*12.0 ± 1.2*
Hairpins with varying barrier height (Fig. 1D)			
Hairpin name	**$\ln(\tau_{1/2})$ (s)**	**$\Delta G^{\ddagger}_{1/2}$ (kJ/mol)**	**$F_{1/2}$ (pN)**
20H21/T4	-2.2 ± 0.6		10.3 ± 0.7
		20.7 ± 0.5	*10.8 ± 0.5*
20H23/T4	-1.8 ± 0.5		10.8 ± 0.8
		22.9 ± 0.4	*11.1 ± 0.5*
20TS10/T4	-0.9 ± 0.3		10.9 ± 0.6
		24.0 ± 0.3	*11.4 ± 0.5*
20H26/T4	-0.1 ± 0.3		11.2 ± 1
		26.4 ± 0.3	*11.7 ± 0.5*
20H32/T4	1.5 ± 0.3		11.0 ± 0.8
		32.1 ± 0.3	*12.3 ± 0.5*
20H38/T4	4.2 ± 0.3		13.4 ± 0.7
		37.7 ± 0.3	*13.4 ± 0.5*
Hairpins with varying mismatch location (Fig. 2C)			
Hairpin name	**$\Delta x_{f\text{-}u}$ (nm)**	**$\Delta x_{f\text{-}i}$ (nm)**	**$\Delta x_{i\text{-}u}$ (nm)**
20R55/T4	18.1 ± 0.3	—	—
	17.9 ± 0.7		
20M04_TT/T4	17.5 ± 0.5	4.1 ± 0.7	13.4 ± 0.6
	18.9 ± 0.7	*3.9 ± 0.4*	*15.0 ± 0.6*
20M06_TT/T4	17.9 ± 0.3	4.5 ± 0.5	13.4 ± 0.4
	18.6 ± 0.6	*4.9 ± 0.4*	*13.7 ± 0.6*
20M07_TT/T4	19.7 ± 0.3	6.0 ± 0.5	13.8 ± 0.5
	18.6 ± 0.6	*5.8 ± 0.4*	*12.8 ± 0.5*
20M10_TT/T4	19.0 ± 0.6	7.3 ± 1.2	11.7 ± 1
	18.5 ± 0.5	*8.8 ± 0.5*	*9.7 ± 0.5*
20M12_TT/T4	18.8 ± 0.6	11 ± 0.9	8.2 ± 0.9
	18.5 ± 0.6	*10.1 ± 0.5*	*8.4 ± 0.5*
20M15_TT/T4	18.8 ± 0.6	14.7 ± 0.8	4.2 ± 0.8
	18.2 ± 0.6	*12.5 ± 0.7*	*5.7 ± 0.6*
20M16_TT/T4	18.2 ± 0.3	14.9 ± 0.9	3.2 ± 0.9
	18.1 ± 0.6	*13.1 ± 0.7*	*5.0 ± 0.7*

Hairpins with varying mismatch identity			
Hairpin name	**$\Delta x_{f\text{-}u}$ (nm)**	**$\Delta x_{f\text{-}i}$ (nm)**	**$\Delta x_{i\text{-}u}$ (nm)**
20M06_AA/T4	19.4 ± 1.1	4.9 ± 0.6	14.5 ± 0.7
	18.7 ± 0.6	*4.6 ± 0.4*	*14.1 ± 0.5*
20M06_GT/T4	19.7 ± 0.6	4.9 ± 0.5	14.9 ± 0.5
	18.7 ± 0.6	*4.6 ± 0.4*	*14.1 ± 0.5*

References

S1. M. T. Woodside *et al*., *Proc. Natl. Acad. Sci. USA* **103**, 6190 (2006).
S2. J. W. Shaevitz, E. A. Abbondanzieri, R. Landick, S. M. Block, *Nature* **426**, 684 (2003).
S3. W. J. Greenleaf, M. T. Woodside, E. A. Abbondanzieri, S. M. Block, *Phys. Rev. Lett*. **95**, 208102 (2005).
S4. K. C. Neuman, S. M. Block, *Rev. Sci. Instr.* **75**, 2787 (2004).
S5. M. Zuker, *Nucl. Acids Res.* **31**, 3406-15 (2003).
S6. J. F. Marko, E. D. Siggia, *Macromolecules* **28**, 209-12 (1995).
S7. M. Peyrard, A. R. Bishop, *Phys. Rev. Lett*. **62**, 2755 (1989).
S8. P. A. Jansson, ed. *Deconvolution of Images and Spectra*, *2nd ed* (Academic Press, New York, 1997).