Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements

Amar Nath Gupta[1][†]**, Abhilash Vincent**[1][†]**, Krishna Neupane**[1]**, Hao Yu**[1]**, Feng Wang**[2]
and Michael T. Woodside[1,2]★

Free-energy-landscape formalisms provide the fundamental conceptual framework for physical descriptions of how proteins and nucleic acids fold into specific three-dimensional structures[1,2]**. Although folding landscapes are difficult to measure experimentally, recent theoretical work by Hummer and Szabo**[3] **has shown that landscape profiles can be reconstructed from non-equilibrium single-molecule force spectroscopy measurements using an extension of the Jarzynski equality**[4]**. This method has been applied to simulations**[5,6] **and experiments**[7,8] **but never validated experimentally. We tested it using force–extension measurements on DNA hairpins with distinct, sequence-dependent folding landscapes. Quantitative agreement was found between the landscape profiles obtained from the non-equilibrium reconstruction and those from equilibrium probability distributions**[9]**. We also tested the method on a riboswitch aptamer with three partially folded intermediate states, successfully reconstructing the landscape but finding some states difficult to resolve owing to low occupancy or overlap of the potential wells. These measurements validate the landscape-reconstruction method and provide a new test of non-equilibrium work relations.**

Folding-landscape formalisms have broad applications in biophysics, from improving predictive structural models and protein engineering[10,11] to providing crucial insights into biomolecular structure, dynamics, function and disease[12,13]. Specific characteristics of folding landscapes, such as the roughness of the free-energy surface[14] or the properties of partially folded intermediate[15] and transition states[16], including how they are altered by temperature changes, solvent substitutions or mutations[17], have been widely studied by experiment and theory. However, it has proven remarkably challenging to go beyond such isolated features and measure the entire profile of the free-energy landscape along the reaction coordinate.

Single-molecule force spectroscopy provides a unique window into folding reactions because of its ability to measure properties of the free-energy landscape. In single-molecule force spectroscopy, a single molecule is held under mechanical tension by a spring-like force probe such as an optical trap or atomic force microscope, and the end-to-end extension of the molecule is recorded while the molecule folds/unfolds under the influence of the denaturing force (Fig. 1a). By this means, the folding energies and rates, partially folded intermediates and similar characteristics may be explored[18]. Measurements can be made either in the equilibrium regime (for example, by maintaining a constant force using an active[19]
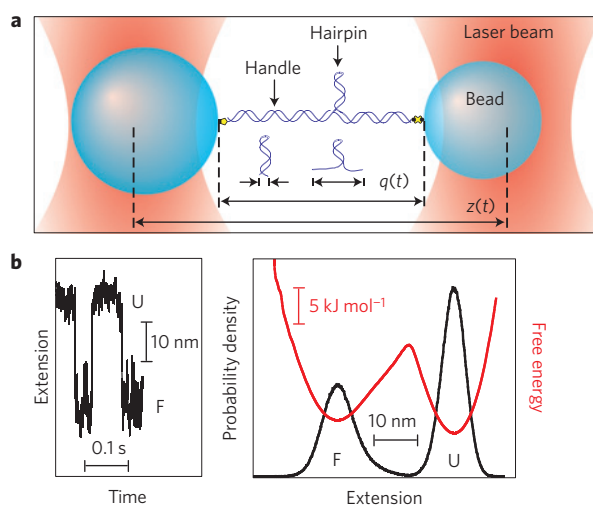

Figure 1 | Optical-trapping measurements. a, Schematic representation of hairpin construct in optical trap. $z(t)$ represents the time-dependent trap separation, $q(t)$ the molecular extension. **b**, Measurements of extension at constant force can be used to generate extension probability distributions and hence the energy-landscape profile. Illustrated here for hairpin A.

or passive[20] force clamp) or in the non-equilibrium regime (for example, by ramping the force rapidly[7]). A method for determining the free-energy profile using equilibrium measurements (Fig. 1b) was recently validated through quantitative comparisons of the landscape profiles measured for different DNA hairpin sequences to a parameter-free model of the sequence-dependent landscapes[9]. This method is accurate and has also been applied to measure protein-folding landscapes[21], but it is technically demanding, requiring very high spatiotemporal resolution and stability.

Equilibrium thermodynamic properties can also be extracted from non-equilibrium measurements such as force-ramp experiments. Although work is irreversibly dissipated in such measurements, the equilibrium free energy can be reconstructed from the distribution of the non-equilibrium work using fluctuation theorems such as the Jarzynski equality[4] or the Crooks fluctuation theorem[22]. These methods have been tested through nanomechanical measurements of unfolding transitions in single RNA hairpins[7,23] and applied widely[24]. Based on an extension of the Jarzynski equality, Hummer and Szabo proposed an approach to

[1]Department of Physics, University of Alberta, 11322-89 Ave, Edmonton AB, T6G 2G7, Canada, [2]National Institute for Nanotechnology, 11421 Saskatchewan Dr, Edmonton AB, T6G 2M9, Canada. [†]These authors contributed equally to this work. *e-mail: Michael.woodside@nrc-cnrc.gc.ca.
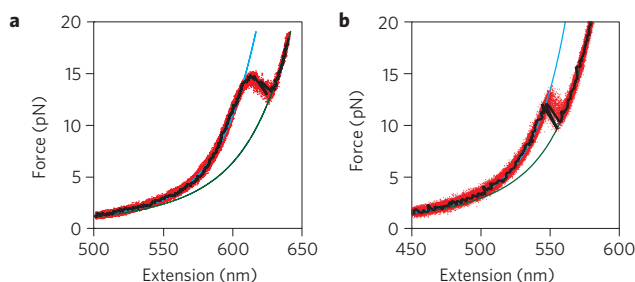
Figure 2 | FECs. a, FECs for hairpin A (99 curves) with WLC fits to the folded (blue line) and unfolded (green line) states. **b**, FECs for hairpin B (266 curves) with WLC fits.
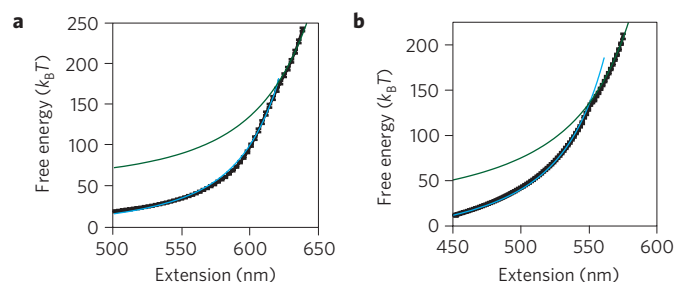

Figure 3 | Free-energy-landscape reconstruction at zero force. a, Energy landscape of hairpin A reconstructed from measurements in Fig. 2a. WLC energies for folded (blue) and unfolded (green) states extrapolated to zero extension reveal the free energy of folding. **b**, Energy landscape for hairpin B reconstructed from Fig. 2b, with WLC fits. All error bars: s.e.m.

reconstructing the free-energy landscape from non-equilibrium force-ramp measurements[3]. The Jarzynski equality relates the free-energy change of a system at a control parameter $z(t)$, $\Delta G_0(z)$, to the non-equilibrium work done, $W(z)$: $\exp(-\Delta G_0/k_BT) = \langle\exp(-W/k_BT)\rangle$, where k_B is the Boltzmann constant, T the temperature, and the average is over all trajectories of z. By converting this expression to a relation describing how a molecular free-energy surface depends on a fluctuating molecular extension, q, the equilibrium profile of the free-energy landscape for a folding transition, $G_0(q)$, was expressed in terms of position histograms as[3]

$$\exp(-G_0(q)/k_BT) = \Big\langle \delta(q - q(t))$$

$$\times \exp\left\{\left[-\int F\,\mathrm{d}q + V(q(0), z(0))\right]/k_BT\right\}\Big\rangle$$

Here $z(t)$ is the position of the force probe, $V(q)$ is the force-probe Hamiltonian, F is the pulling force and δ is the Dirac delta function.

This approach has been applied in numerous studies to reconstruct the free-energy profile, in the context of both simulations[5,6] and experiments[7,8]; however, it has never been validated by an experimental comparison to other, established methods for determining the folding-landscape profile. To validate the method, we measured force–extension curves (FECs) for DNA hairpins with different, sequence-dependent landscapes that were previously known from equilibrium measurements[9], and compared the non-equilibrium and equilibrium results. Single DNA hairpins (consisting of a simple stem–loop structure) attached to kilobase-long handles of double-stranded DNA were bound to polystyrene beads held in dual-beam optical tweezers (Fig. 1a). Sets of FECs unfolding the hairpins were measured by moving the traps apart at a constant velocity and recording the force on the molecule as a function of extension until the hairpins unfolded. Two different hairpin stem sequences producing different energy-landscape profiles were chosen to test the non-equilibrium landscape reconstruction: hairpin A, with a 30-base-pair (bp) stem sequence producing a barrier close to the unfolded state[9,25], and hairpin B, with a 20-bp stem sequence producing an energy barrier close to the folded state[9].

A total of 1,630 FECs were measured for hairpin A and 2,918 for hairpin B, in four different sets for each hairpin. Within each set, FECs were aligned to remove the small instrumental drift (typically a few nanometres or less) that occurred during the experiments. Individual FECs, shown in black atop the aggregated data from a full set for each hairpin (Fig. 2), show the expected behaviour: force increases nonlinearly with extension, owing to the handle elasticity[26], until the hairpin unfolds cooperatively in a two-state manner, creating characteristic 'sawtooth' patterns as the extension suddenly increases[27]. The FECs are well fitted by an extensible worm-like chain (WLC) polymer model[28] using two chains in series:

one for the double-stranded DNA of the handles and one for the single-stranded DNA from the unfolded hairpin.

The unperturbed (zero-force) free-energy profile, $G_0(q)$, was calculated from the FECs using the weighted-histogram method[3] (Fig. 3, black). $G_0(q)$ represents the energy profile for the entire molecule, including the handles as well as the hairpin. It is dominated by the stretching energy of the handles, as seen by fitting $G_0(q)$ to the integral of the WLC fits from Fig. 2 (blue, folded state; green, unfolded state). The kink in the profile near the crossing point of the two fits represents the unfolding of the hairpins, and hence is the region of interest. As a consistency check, we note that the energy differences between the two WLC fits at $q = 0$, respectively $60 \pm 2\,k_BT$ and $36 \pm 2\,k_BT$ for hairpins A and B, match the values for the unfolding free energy obtained from constant-force measurements ($65 \pm 5\,k_BT$ and $35 \pm 2\,k_BT$, respectively[9,25]) as well as those calculated from modelling[9,25] ($62 \pm 4\,k_BT$ and $38 \pm 3\,k_BT$).

The profile $G_0(q)$ is the fundamental result of the landscape reconstruction, but to validate the method we compared $G_0(q)$ to the free-energy profile, $G_{eqm}(q)$, determined from the inverse Boltzmann transform of extension histograms measured at constant force[25]: $G_{eqm}(q) = -k_BT\ln(P(q))$, where $P(q)$ is the probability distribution of the molecular extension. $P(q)$ was found by measuring the extension of the hairpin constructs at high bandwidth (50 kHz) near $F = F_{1/2}$, the force at which the hairpin was poised equally between folded and unfolded states. The force was kept constant using a passive force clamp[20] to avoid artefacts from active feedback loop closure. $G_0(q)$ was transformed into a constant-force profile, $G_F(q)$, by subtracting the work done by the constant force F_0: $G_F(q) = G_0(q) - F_0q$.

The results from each of the four sets of FECs for each hairpin are shown in Fig. 4 (open circles) along with $G_{eqm}(q)$ (black lines). Error bars represent the standard error in the reconstructions, estimated by bootstrap analysis. Good agreement is found between the non-equilibrium and equilibrium reconstructions: for example, the barrier locations agree within one histogram bin, barrier heights are typically within one standard error of the equilibrium result and the general shapes of the profiles are very similar. The principal difference is in the shapes of the potential wells of the folded and unfolded states: the wells in the non-equilibrium reconstruction on average have slightly higher curvature, probably owing to the fact that the effective trap stiffness used for the force-ramp measurements was higher than that for the equilibrium measurements.

These results validate the non-equilibrium method. We note that the four sets of data for hairpin A were measured at four different pulling rates (10, 20, 100 and 333 nm s^{-1}, with 99, 100, 138 and 1,293 FECs respectively), producing dissipated energies of 1.1 ± 0.7, 1.4 ± 0.7, 3.2 ± 0.7 and $4.9 \pm 0.3\,k_BT$, respectively.
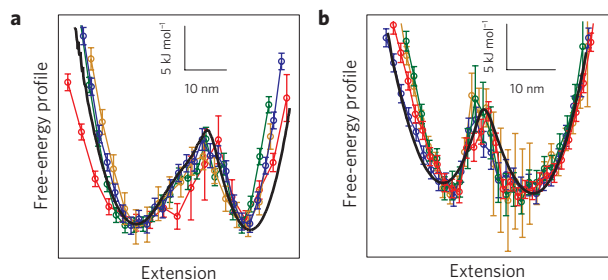
Figure 4 | Comparison of equilibrium and non-equilibrium reconstructions. a,b, The energy landscapes near $F_{1/2}$ for hairpin A (**a**) and hairpin B (**b**) reconstructed from FECs (open circles) agree with those reconstructed from constant-force measurement (black lines). The pulling rate for hairpin B was 500 nm s^{-1} (all curves); for hairpin A, it was 10 nm s^{-1} (green), 20 nm s^{-1} (yellow), 100 nm s^{-1} (blue) and 333 nm s^{-1} (red). Error bars: s.e.m. Standard error in equilibrium profiles at barrier peak: 0.3 kJ mol^{-1} (hairpin A), 0.4 kJ mol^{-1} (hairpin B).

The FECs at 20 nm s^{-1} are shown in Fig. 2a; representative FECs at the other speeds are shown in Supplementary Information. All FECs for hairpin B were measured at 500 nm s^{-1}, resulting in $5.5 \pm 0.3 \, k_B T$ of dissipated energy. Little difference was observed in the reconstructions at different rates, although the errors were larger for the fastest rates, as would be expected owing to the need to sample exponentially more curves when further away from equilibrium[29]. Significantly, a large number of FECs is not required to obtain a reasonable reconstruction of the landscape profile, making this method more experimentally practical in many cases than the equilibrium method.

A key feature of these reconstructions is that they include the effects of the handles, which smooth out the sharp landscape features. To obtain the landscape of the hairpin alone, the handle effects would need to be removed by deconvolution[9]. This effect is relevant to almost all single-molecule force spectroscopy data, and complicates the interpretation because the smoothing typically changes both the location and height of the energy barrier. It is especially important for analysing data from tandem-repeat protein constructs, where the 'handles' consist of unfolded protein: the handle length (and hence smoothing) changes as each subunit unfolds. In such experiments, the sequential unfolding transitions, which each involve different-length unfolded polypeptide handles, should be analysed separately[8].

Whereas two-state hairpins are a good model system for validating the method, many molecules unfold through intermediate states. To test whether the Hummer–Szabo method can capture multiple intermediates, we measured FECs of the *add* adenine riboswitch aptamer, which has three partially folded intermediates similar to a previously studied adenine riboswitch aptamer[30]. 700 FECs from a single measurement set (Fig. 5a) were analysed as for the hairpins to obtain $G_0(q)$ (Fig. 5b). The resultant zero-force landscape was then tilted by force to compare with $G_{eqm}(q)$ (Fig. 5c) as determined from constant-force data (Fig. 5d). Once again, $G_0(q)$ matches $G_{eqm}(q)$ reasonably well. Notably, although there seem to be only two potential wells, the high-extension well is much broader than for the two-state hairpins (Fig. 4): in fact, this well arises from three distinct states, the unfolded state and two intermediates (indicated by red arrows), which overlap partially in extension and force. These states are visible in the filtered constant-force data (Fig. 5d, black, delineated by red lines), although only the two intermediates are highly populated at this force. One more intermediate can also be seen in the filtered constant-force data (Fig. 5d, blue line), but it is too rarely populated to appear in the equilibrium reconstruction. Since this state is not present in any of the FECs, it is also absent from
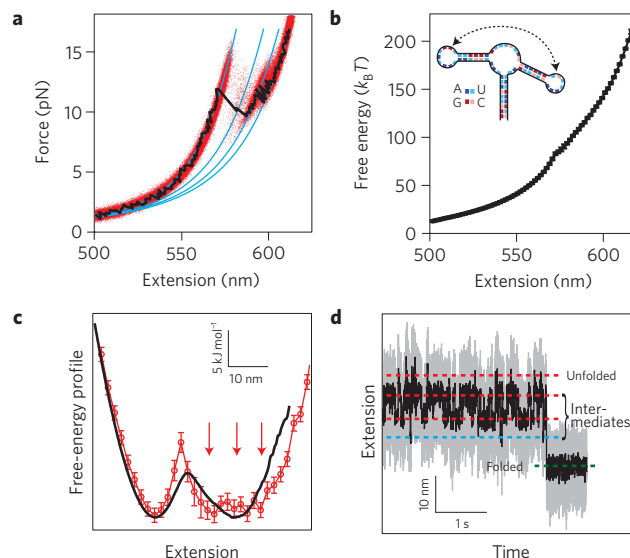

Figure 5 | Riboswitch-aptamer landscape reconstruction. a, 700 FECs for the *add* adenine riboswitch aptamer show multiple intermediate states (blue lines: WLC fits). **b**, Free-energy profile at zero force. Error bars: s.e.m. Inset: Aptamer secondary structure. **c**, The landscape at constant force reconstructed from FECs (open circles) agrees with that reconstructed from constant-force measurements (black line). The right-hand potential well contains contributions from three states (arrows) that are not well resolved. Error bars: s.e.m. **d**, The aptamer extension at constant force (grey), median-filtered at 2.5 ms (black), shows three highly occupied states and two low-occupancy states (dotted lines). Blue line: intermediate not seen in the FECs.

the non-equilibrium reconstruction. These results demonstrate the ability to reconstruct multistate landscapes, while also indicating some of the technical challenges involved in interpreting the reconstructions.

Methods

DNA hairpins attached to double-stranded DNA handles were prepared as described previously[25]. The resulting constructs contained a biotin-labelled 798-bp handle on the 3' end of the hairpin, a digoxigenin-labelled 1,158-bp handle on the 5' end of the hairpin and abasic sites separating the hairpin from each handle. The constructs were incubated with 600-nm- and 820-nm-diameter polystyrene beads labelled with avidin DN (Vector Labs) and anti-digoxigenin (Roche), respectively, to create dumbbells. Dumbbells were diluted to ~2 pM in measuring buffer (50 mM MOPS at pH 7.5, 200 mM KCl and oxygen scavenging system: 40 U ml^{-1} glucose oxidase, 185 U ml^{-1} catalase and 8.3 mg ml^{-1} glucose) and inserted into a sample chamber on a clean microscope slide in the optical trap. Hairpin A has the same sequence as 30R50/T4 (ref. 25) and hairpin B has the same sequence as 20TS06/T4 (ref. 9).

Measurements were made using a custom-built dual-beam optical trap similar to previous descriptions[20]. Two 1,064 nm laser beams with independent position and intensity control were used to apply force to the hairpins, and the resulting changes in molecular extension were measured from the deflection of two 633 nm laser beams scattered onto position-sensitive diodes. The trap stiffness used was 0.3 and 0.9 pN nm^{-1}. FECs were measured by moving the traps apart in 1.0 nm steps at different rates (10–500 nm s^{-1}) using electro-optic deflectors (Conoptics) controlled by custom LabVIEW software (National Instruments). At each step, data were sampled at 20 kHz, filtered online at 10 kHz by an eight-pole Bessel filter and averaged. Offline analysis was done using Igor Pro custom software (Wave metrics).

Free-energy profiles were reconstructed following a previously described procedure[3]. Briefly, we assumed a time-dependent Hamiltonian of the single-molecule pulling system of the form $H(x,t) = H_0(x) + V(x,t)$, where $H_0(x)$ is the Hamiltonian of the unperturbed system and $V(x,t)$ is the time-dependent perturbation from the trap. The perturbation was described in terms of the molecular extension $q(x)$, effective trap stiffness k_s and centre position $z(t) = z_0 + vt$ of the trap moving at velocity v as $V(x,t) = V[q(x) - z(t)] = 1/2 k_s [q(x) - z(t)]^2$.

The external work along each non-equilibrium trajectory at time t was then found from

$$W_t = \int_0^t \frac{\partial H(x(t'),t')}{\partial t} \, dt' = k_s v \left[vt^2/2 + z_0 t - \int_0^t z_{t'} \, dt' \right]$$

For each FEC, indexed by k, we determined the trap position z_{ik} at discrete time t_i and then calculated the external work W_{ik} by numerical integration of the data using

$$W_{ik} = k_s v \left[0.5 \, vt_i^2 + z_0 \, t_i - \sum_{j=1}^{i} (t_j - t_{j-1})(q_{jk} + q_{j-1,k})/2 \right]$$

To obtain an optimal estimate of the free-energy surface, we averaged over many time slices and repeated trajectories. For each time slice t, an ensemble of positions z_t and corresponding W_t values was obtained. The extensions were binned with respect to the time slices and the corresponding histogram values were weighted by $\exp(-\beta W_t)$, where $\beta = 1/k_B T$, leading to the 'weighted-histogram' calculation of the unperturbed molecular free energy:

$$G_0(q) = -\beta^{-1} \ln \frac{\sum_t \frac{\langle \delta(q-q_t)\exp(-\beta W_t)\rangle}{\langle \exp(-\beta W_t)\rangle}}{\sum_t \frac{\exp[-\beta V(q,t)]}{\langle \exp(-\beta W_t)\rangle}}$$

Bin sizes (1.5–3.5 nm) were chosen to ensure sufficient data within each bin to define the distribution of work for that bin reliably[6]. Errors in $G_0(q)$ were determined by bootstrap analysis. Since $G_0(q)$ represents the free-energy profile at $F = 0$, to compare with the energy profile measured at constant force (F_0), an additional energy given by $F_0 q$ was subtracted from the non-equilibrium energy reconstruction.

References

1. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **48,** 545–600 (1997).
2. Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl Acad. Sci. USA* **84,** 7524–7528 (1987).
3. Hummer, G. & Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl Acad. Sci. USA* **98,** 3658–3661 (2001).
4. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78,** 2690 (1997).
5. Park, S. & Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **120,** 5946–5961 (2004).
6. Minh, D. D. L. Free-energy reconstruction from experiments performed under different biasing programs. *Phys. Rev. E* **74,** 061120 (2006).
7. Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I. J. & Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* **296,** 1832–1835 (2002).
8. Harris, N. C., Song, Y. & Kiang, C-H. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality. *Phys. Rev. Lett.* **99,** 068101 (2007).
9. Woodside, M. T. *et al.* Direct measurement of the full sequence-dependent folding landscape of a nucleic acid. *Science* **314,** 1001–1004 (2006).
10. Petrey, D. & Honig, B. Protein structure prediction: Inroads to biology. *Mol. Cell.* **20,** 811–819 (2005).
11. Bradley, P., Misura, K. M. S. & Baker, D. Toward high-resolution *de novo* structure prediction for small proteins. *Science* **309,** 1868–1871 (2005).
12. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, 1999).
13. Dobson, C. M. Protein folding and misfolding. *Nature* **426,** 884–890 (2003).
14. Hyeon, C. & Thirumalai, D. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proc. Natl Acad. Sci. USA* **100,** 10249–10253 (2003).
15. Kim, P. S. & Baldwin, R. L. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59,** 631–660 (1990).
16. Daggett, V. & Fersht, A. R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28,** 18–25 (2003).
17. Buchner, J. & Kiefhaber, T. *Protein Folding Handbook* (Wiley-VCH, 2005).
18. Woodside, M. T., García-García, C. & Block, S. M. Folding and unfolding single RNA molecules under tension. *Curr. Opin. Chem. Biol.* **12,** 640–646 (2008).
19. Lang, M. J., Asbury, C. L., Shaevitz, J. W. & Block, S. M. An automated two-dimensional optical force clamp for single molecule studies. *Biophys. J.* **83,** 491–501 (2002).
20. Greenleaf, W. J., Woodside, M. T., Abbondanzieri, E. A. & Block, S. M. Passive all-optical force clamp for high-resolution laser trapping. *Phys. Rev. Lett.* **95,** 208102 (2005).
21. Gebhardt, J. C. M., Bornschlögl, T. & Rief, M. Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl Acad. Sci. USA* **107,** 2013–2018 (2010).
22. Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **60,** 2721 (1999).
23. Collin, D. *et al.* Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* **437,** 231–234 (2005).
24. Ritort, F. Nonequilibrium fluctuations in small systems: From physics to biology. *Adv. Chem. Phys.* **137,** 31–123 (2008).
25. Woodside, M. T. *et al.* Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl Acad. Sci. USA* **103,** 6190–6195 (2006).
26. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271,** 795–798 (1996).
27. Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. J. & Bustamante, C. Reversible unfolding of single RNA molecules by mechanical force. *Science* **292,** 733–737 (2001).
28. Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. Stretching DNA with optical tweezers. *J. Biophys.* **72,** 1335–1346 (1997).
29. Mossa, A., Lorenzo, S. d., Huguet, J. M. & Ritort, F. Measurement of work in single-molecule pulling experiments. *J. Chem. Phys.* **130,** 234116 (2009).
30. Greenleaf, W. J., Frieda, K. L., Foster, D. A. N., Woodside, M. T. & Block, S. M. Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **319,** 630–633 (2008).

Author contributions

M.T.W. designed the experiment. F.W. contributed materials. K.N., H.Y. and M.T.W. made the measurements. A.N.G., A.V. and K.N. analysed the data. A.N.G., A.V. and M.T.W. wrote the paper.

Additional information

The authors declare no competing financial interests. Supplementary information accompanies this paper on www.nature.com/naturephysics. Reprints and permissions information is available online at http://www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.T.W.