Reconstructing Folding Energy Landscape Profiles from Nonequilibrium Pulling Curves with an Inverse Weierstrass Integral Transform

Megan C. Engel,[1] Dustin B. Ritchie,[1] Daniel A. N. Foster,[1] Kevin S. D. Beach,[1,2] and Michael T. Woodside[1,3,*]

[1]*Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2E1 Canada*
[2]*Department of Physics and Astronomy, University of Mississippi, University, Mississippi 38677 USA*
[3]*National Institute for Nanotechnology, National Research Council, Edmonton, Alberta, T6G 2M9 Canada*

The energy landscapes that drive structure formation in biopolymers are difficult to measure. Here we validate experimentally a novel method to reconstruct landscape profiles from single-molecule pulling curves using an inverse Weierstrass transform (IWT) of the Jarzysnki free-energy integral. The method was applied to unfolding measurements of a DNA hairpin, replicating the results found by the more-established weighted histogram (WHAM) and inverse Boltzmann methods. Applying both WHAM and IWT methods to reconstruct the folding landscape for a RNA pseudoknot having a stiff energy barrier, we found that landscape features with sharper curvature than the force probe stiffness could not be recovered with the IWT method. The IWT method is thus best for analyzing data from stiff force probes such as atomic force microscopes.

Energy landscape theory provides the fundamental framework for the modern description of structural self-assembly, or "folding," of biopolymers like proteins and nucleic acids [1,2]. In this picture, folding is viewed as a diffusive process on a hypersurface describing the free energy as a function of the conformational degrees of freedom. The landscape encapsulates the critical properties governing folding, including energy changes, barriers, intermediates, roughness and internal friction, frustration, alternative pathways connecting states, effects of sequence mutations, and folding reaction rates, amongst other things. Mapping and understanding energy landscapes is thus a key goal in folding research.

Despite the importance of energy landscapes for understanding folding and the wealth of existing theoretical and computational work in landscape theory, few methods exist for reconstructing energy landscapes quantitatively from experimental data. One platform that has proven effective in this regard is single-molecule force spectroscopy (SMFS), whereby a denaturing tension is applied to an individual molecule by a force probe and the extension of the molecule is measured as its structure changes in response to the load [3,4]. Since the molecular extension represents a one-dimensional (1D) reaction coordinate for the folding in SMFS, the full landscape is projected onto a 1D profile [5]. Such 1D projections can nevertheless successfully capture the important features of the landscape [6] and reproduce the folding dynamics [7].

Several methods have been developed for reconstructing the energy landscape profile from different modalities of SMFS [8], using measurements of "hopping" between structural states in equilibrium at constant or near-constant force [9–11] or nonequilibrium measurements where the force is changed rapidly [12–14]. Of particular interest are methods applicable to force-ramp experiments, wherein force is ramped up gradually to unfold a molecule, as they constitute one of the most commonly applied SMFS modalities. By extending the Jarzysnki equality relating the free energy change to the irreversible work done [15], Hummer and Szabo [12] proposed one such method for landscape reconstruction. In their approach, the free-energy profile as a function of molecular extension, q, is computed from Jarzynski estimates of the equilibrium energy calculated within different time windows and combined using the weighted-histogram analysis method (WHAM). The free-energy surface $G_0(q)$ at zero force is

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

Here β is the inverse thermal energy, $\delta(q)$ the Dirac δ function, $W_t = \int_0^t (\partial V/\partial z)(\partial z/\partial t') dt'$ is the external work done on the system by the force probe in time t during a particular trial, and $V(q, z) = k[q - z(t)]^2/2$ is the perturbing potential from the probe, with k representing the probe stiffness and $z(t)$ the average probe position at time t. Note that q is the molecular extension whereas z is the experimental control parameter for the reaction.

This approach has been used successfully to reconstruct full energy profiles [7,13,16], but it suffers from the need to obtain sufficient statistics within each time bin to perform a reliable Jarzynski average, which can result in relatively coarse sampling of the landscape profile [13]. In some cases, insufficient data may be available to reconstruct the landscape in some ranges of the molecular extension [16].

Recently, an alternate approach was developed by Hummer and Szabo to address this issue and simplify the landscape reconstruction [17]. Instead of reconstructing $G_0(q)$ directly from the pulling curves, a two-step method is pursued. First, the free energy as a function of the probe position, $A(z)$, is computed from the accumulated work as a function of the probe position, $W(z)$, using the Jarzynski identity [15]: $\beta A(z) = -\langle \exp[-\beta W(z)]\rangle$. Next, the free energy as a function of molecular extension, $G_0(q)$, is recovered from the partition function:

$$\exp[-\beta A(z)] = \int \exp[-\beta G_0(q) - \beta k(q-z)^2/2]dq. \quad (2)$$

Equation (2) is mathematically equivalent to a Weierstrass transformation [18], which is a convolution of a function—in this case the exponential of $G_0(q)$—with a Gaussian, here of variance $1/\beta k$.

Equation (2) can be evaluated as described previously [17], using Laplace's method, under the assumption that the integrand has a strict maximum at the stationary point $q_0 = z - G_0'(q_0)/k$, which is consistent with the requirement that $G_0''(q_0) > k$. In practice, this means that $G_0(q)$ reconstructed in this way will not capture features of the landscape whose curvature is sharper than the spring constant of the force probe. The result, to second order, is

$$G_0\left(q = z - \frac{A'(z)}{k}\right) = A(z) - \frac{A'(z)^2}{k} + \frac{1}{2\beta}\ln\left[1 - \frac{A''(z)}{k}\right]. \quad (3)$$

Here, the derivatives of $A(z)$, taken with respect to z, can be determined via work-weighted trajectory averages:

$$A'(z) = -k\langle\langle q - z\rangle\rangle = \langle\langle F\rangle\rangle \quad (4)$$

and

$$1 - A''(z) = \frac{\beta}{k}(\langle\langle F^2\rangle\rangle - \langle\langle F\rangle\rangle^2), \quad (5)$$

where the double angle brackets $\langle\langle\ldots\rangle\rangle$ indicate an average taken over all trajectories according to $\langle(\ldots)\exp[-\beta W(z)]\rangle/\langle\exp[-\beta W(z)]\rangle$.

Whereas the WHAM approach has been validated by comparison to other methods for landscape reconstruction [13], no such experimental test has yet been done for the inverse Weierstrass transform (IWT) approach, despite its promise for expanding the applicability of experimental landscape reconstruction methods. We performed such a validation by comparing the WHAM and IWT landscape reconstructions from SMFS measurements of two different molecules, a DNA hairpin and a RNA pseudoknot, using optical tweezers. First, single DNA hairpin molecules were prepared attached to kilobase (kb)-long handles of double-stranded (ds) DNA, as described previously [19], and
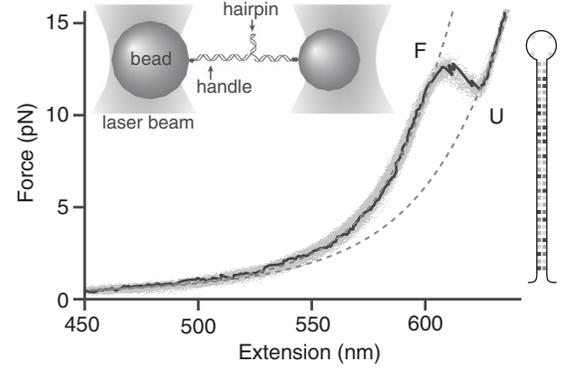

FIG. 1.   Force-extension curves of a DNA hairpin (diagram, right) measured with an optical trap. Dashed lines: wormlike chain fits to the folded (F) and unfolded (U) states. Inset: schematic illustrating the measurement.

bound specifically at each end to functionalized polystyrene beads held in a high-resolution dual-beam optical trap (Fig. 1, inset). The hairpin (sequence 30R50/T4) is known to fold as a two-state system under tension [19]. The two traps were moved apart at a constant rate, with pulling speeds ranging from 10 to 300 nm/s, and the force was measured as a function of the molecular extension, generating force-extension curves (FECs) as in Fig. 1.

The landscape profile $G_0(q)$ for this hairpin was first calculated via the WHAM method using Eq. (1). For comparison to the profile derived from an inverse Boltzmann transform of the equilibrium extension distribution in a constant-force experiment (Fig. 2, dotted line) [9], the solution was tilted to the force $F_{1/2}$, at which the folded- and unfolded-state wells were of equal depth, as described previously [13]. The result, averaging over 16 data sets containing 1641 FECs, is shown in Fig. 2 (black).
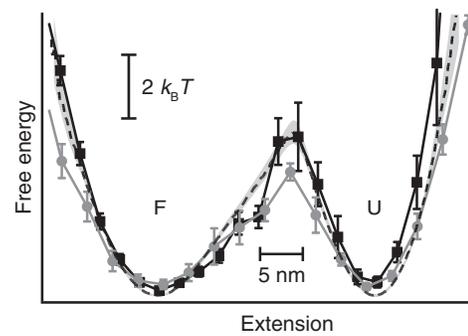

FIG. 2.   Profile of the free-energy landscape along the molecular extension of the DNA hairpin. The profile calculated from the force-extension curves using WHAM (black) agrees well with the profile found from constant-force measurements via an inverse Boltzmann transform (dashed line). The profile found using the IWT approach (gray) agrees reasonably well with the other two profiles, but underestimates the barrier height slightly owing to incomplete defiltering of smoothing caused by the force probe. Error bars and shaded gray region show standard errors.

Next, $G_0(q)$ was calculated from 824 FECs in 8 data sets with the IWT method, using the work-weighted averages according to Eqs. (3)–(5), and again tilted to $F_{1/2}$ (Fig. 2, gray). All three landscapes agree reasonably well: the positions of the potential wells and the barrier are the same to within 1 nm for each of the reconstructions, and the energies differ by less than $\sim 1 k_B T$ everywhere along the profile, including at the barrier. The IWT technique is thus validated as a viable alternative to the more-established reconstruction methods.

Several features of these reconstructions are notable. First, there is a relatively small ($\leq 1 k_B T$) but systematic underestimate of the barrier height in the IWT reconstruction. This effect likely reflects one of the limitations of the IWT approach, namely, that the Weierstrass transform filters $G_0(q)$ by a Gaussian whose width varies inversely with the probe stiffness. Features in the landscape that are stiffer than the force probe are therefore smoothed out, as discussed in more detail below. Interestingly, the variance in the landscape constructed using the IWT is noticeably lower than that in the WHAM landscape, suggesting that the IWT is a more robust approach requiring fewer curves and avoiding the aforementioned coarse-graining problem of the WHAM reconstructions. This advantage of the IWT approach was noted by Hummer and Szabo [17]; it arises from the fact that each point on the landscape is calculated from a combination of all FECs, whereas in the WHAM approach, each point is the result of only the few FECs that fall into the corresponding extension bin.

To explore the application of landscape reconstruction using the IWT approach in the context of a more complex molecule containing tertiary structure, we also studied a RNA pseudoknot. Pseudoknot structures are widespread in nature, playing functionally diverse roles such as stimulating programmed −1 ribosomal frameshifting (−1 PRF) [20,21] and maintaining telomere length in highly proliferative cells such as stem cells or oncogenic cells [22]. The activity of a pseudoknot is often regulated at the level of structural changes, including mechanical unfolding and conformational equilibria. For example, changes in the conformational equilibrium between competing pseudoknot and hairpin structures cause dyskeratosis congenita [23,24], and stimulatory pseudoknots determine the efficiency of −1 PRF through a mechanism that remains controversial [20,21] but appears to involve mechanical unfolding by the ribosome [25] and conformational plasticity in the pseudoknot [26,27]. Reconstructing quantitative energy landscapes should therefore provide insight into how the dynamics and energetics of such RNA structures relate to their biological function, possibly aiding the development of therapeutics. The mechanical unfolding of pseudoknots has been studied previously with SMFS [26–31], but the full profile of the energy landscape of a pseudoknot has not yet been experimentally reconstructed.

We focused on the pseudoknot stimulating −1 PRF in the sugar cane yellow leaf virus [32]. RNA containing the pseudoknot sequence was transcribed and annealed with single-stranded DNA to create a construct consisting of the RNA pseudoknot [Fig. 3(a), inset] flanked by kb-long RNA-DNA duplex handles, analogous to the hairpin constructs, as described previously [26]. FECs measured at pulling speeds of 100–270 nm/s revealed a much wider range of unfolding forces than for the hairpin [Fig. 3(a)]. The contour length change upon unfolding, found from fitting the FECs to an extensible wormlike chain model [26,33], agreed with the value expected from the structure of the pseudoknot [32], confirming that it was natively folded. Analyzing 1217 FECs from 11 data sets, we reconstructed the landscape profile for the pseudoknot using both the WHAM [Fig. 3(b), black] and IWT [Fig. 3(b), gray] approaches, similar to the hairpin. To our knowledge, this represents the first landscape reconstruction for a pseudoknot.

The result from the WHAM analysis reveals a barrier that is closer to the folded state than the unfolded state, in contrast to the barrier for hairpin unfolding. This difference, indicative of a more "brittle" structure for the pseudoknot
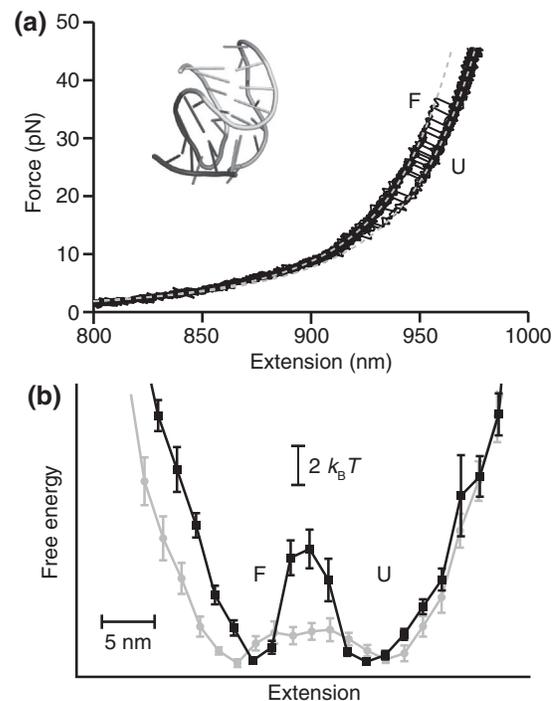

FIG. 3. (a) Force-extension curves of the sugar cane yellow leaf virus pseudoknot. Dashed lines: wormlike chain fits to the folded (*F*) and unfolded (*U*) states. Inset: structure of the pseudoknot. (b) Profiles of the free-energy landscape reconstructed from the FECs. The profile found using WHAM (black) has a barrier close to the folded state. The profile found using IWT (gray) does not agree very well, showing a much lower barrier and broader wells. These profiles constitute the first experimental landscape reconstructions for a RNA pseudoknot.

than the hairpin [8], is also reflected in the broader distribution of unfolding forces [34]; indeed, it is expected for structures like pseudoknots in which unfolding involves shearing of base pairs and tertiary interactions [28,29,35–37], rather than merely base-pair unzipping as in hairpins [9,19]. Notably, the landscape reconstructed via the IWT differed significantly from the WHAM result in this case, failing to recover the sharp barrier in the pseudoknot landscape. Instead, the barrier in the IWT reconstruction was several-fold lower and broader, located further from the folded state, and the potential wells on each side were also broader.

The discrepancies between the WHAM and IWT results for the pseudoknot can be understood in terms of the filtering effect of the probe stiffness on the IWT reconstruction, as mentioned above. Equation (3) assumes that the probe stiffness exceeds the curvature of any features in the landscape. To test whether this assumption holds, we measured the stiffness of the barriers recovered from the WHAM analyses of the hairpin and the pseudoknot using parabolic fits (Fig. 4), and in each case compared the barrier stiffness to the effective stiffness of the force probe (here, the laser traps) used in the measurement. For the hairpin, the barrier had a stiffness of 0.5 ± 0.1 pN/nm [Fig. 4(a), dot-dashed line], just slightly higher than the effective probe stiffness of 0.4 pN/nm. The approximation in the analysis is thus not unreasonable, and the WHAM (black) and IWT (dark gray) reconstructions agree fairly well; the fact that the barrier height is slightly too low in the IWT reconstruction can be attributed to a small effect from the force probe, since the barrier is lowered to the point where it has a curvature comparable to that of the probe potential [Fig. 4(a), dashed line]. For the pseudoknot, on the other hand, the barrier in the WHAM result [Fig. 4(b), black] has a curvature of 2.1 pN/nm [Fig. 4(b), dashed line], almost an order of magnitude higher than the effective stiffness of the force probe in these measurements, 0.26 pN/nm [Fig. 4(b), dotted line]. The approximation in the IWT analysis thus breaks down, and the suppression of the barrier by the compliance of the probe changes the result significantly. This filtering effect is illustrated clearly by repeating the IWT analysis on hairpin measurements (893 FECs in 9 data sets) made at lower trap stiffness, 0.24 pN/nm [Fig. 4(a), light gray]: the barrier is suppressed to the point where it has curvature comparable to that of the force probe [Fig. 4(a), dotted line], opening up a more substantial disagreement between the WHAM and IWT results.

Defiltering, as done by the IWT, is, in principle, an ill-posed problem. In practice, however, we found that the IWT approach yielded stable, reproducible answers, despite the presence of errors from noise and finite sampling, giving confidence that it can be applied reliably to experimental data. Nevertheless, fine-scale details such as sharp barriers (as above) or local
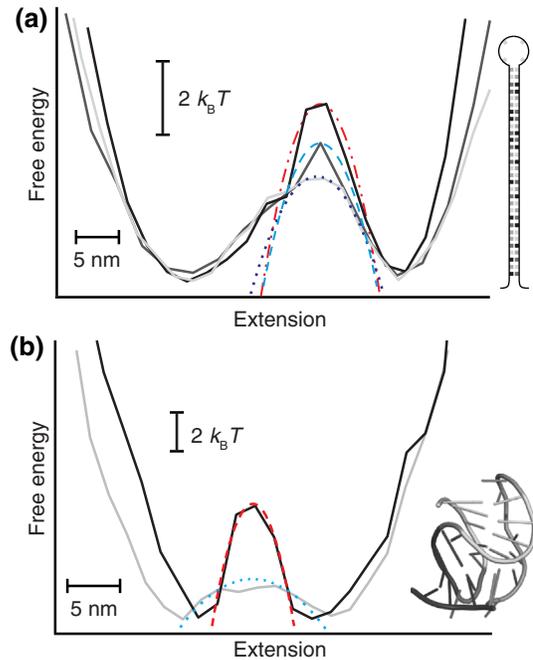

FIG. 4 (color online). Effect of the probe stiffness on the landscape reconstruction using IWT. (a) For the hairpin, the stiffness of the probe (dashed line) is only slightly lower than the stiffness of the barrier (dot-dashed line) found by fitting the WHAM profile (black). The IWT reconstruction (dark gray) is therefore largely successful, although the probe compliance reduces the barrier height slightly. For measurements made at half the barrier stiffness (light gray), the barrier is suppressed further, matching the probe stiffness (dotted line). (b) For the pseudoknot, the probe stiffness (dotted line) is much lower than the stiffness of the barrier (dashed line) from the WHAM reconstruction (black), and the IWT reconstruction (gray) is thus not successful.

roughness in the landscape [38,39] typically cannot be recovered. We note that the problem of resolving fine details in the landscapes of biomolecules is ubiquitous across different reconstruction approaches, whether due to Gaussian filtering by a probe as above, coarse binning [8], or inadequate sampling [40].

In addition to presenting the first reconstruction of the landscape for a pseudoknot, these results provide the first experimental validation of the IWT method, demonstrating its utility as a complement to other landscape reconstruction techniques and its practical limitations in experimental applications. Our results also underline an important advantage of this approach: it reduces the error in the reconstruction, or equivalently requires fewer curves to reconstruct the landscape, as compared to WHAM reconstructions. This feature of the method extends the range of systems whose landscapes can be reconstructed, for example allowing reconstructions even if only a small number of FECs can be measured for a particular structure [41]. The need to use a stiff force probe may limit the potential applications of the IWT

approach for measurements made with optical traps, which tend to have relatively low stiffness [42], although high-stiffness trapping applications have been demonstrated [43]. However, it renders the IWT approach ideal for analyzing measurements made using atomic force microscopes (AFMs), which have much higher probe stiffness but for which the WHAM approach can be problematic [16]: even molecules with very stiff barriers will be amenable to study by AFM. As AFMs are widely used to probe folding, especially in proteins, we expect the IWT approach to become an important tool in ongoing efforts to make the reconstruction of energy landscapes a more routine and widely applied feature of folding studies.

*michael.woodside@ualberta.ca

[1] J. N. Onuchic and P. G. Wolynes, Curr. Opin. Struct. Biol. **14**, 70 (2004).

[2] K. A. Dill and J. L. MacCallum, Science **338**, 1042 (2012).

[3] M. T. Woodside, C. Garcia-Garcia, and S. M. Block, Curr. Opin. Chem. Biol. **12**, 640 (2008).

[4] G. Zoldak and M. Rief, Curr. Opin. Struct. Biol. **23**, 48 (2013).

[5] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **107**, 1088 (2010).

[6] O. K. Dudko, T. G. W. Graham, and R. B. Best, Phys. Rev. Lett. **107**, 208301 (2011).

[7] H. Yu, A. N. Gupta, X. Liu, K. Neupane, A. M. Brigley, I. Sosova, and M. T. Woodside, Proc. Natl. Acad. Sci. U.S.A. **109**, 14452 (2012).

[8] M. T. Woodside and S. M. Block, Annu. Rev. Biophys. **43**, 19 (2014).

[9] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block, Science **314**, 1001 (2006).

[10] J. C. M. Gebhardt, T. Bornschloegla, and M. Rief, Proc. Natl. Acad. Sci. U.S.A. **107**, 2013 (2010).

[11] M. de Messieres, B. Brawn-Cinani, and A. La Porta, Biophys. J. **100**, 2736 (2011).

[12] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).

[13] A. N. Gupta, A. Vincent, K. Neupane, H. Yu, F. Wang, and M. T. Woodside, Nat. Phys. **7**, 631 (2011).

[14] H. Lannon, J. S. Haghpanah, J. K. Montclare, E. Vanden-Eijnden, and J. Brujic, Phys. Rev. Lett. **110**, 128301 (2013).

[15] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[16] N. C. Harris, Y. Song, and C.-H. Kiang, Phys. Rev. Lett. **99**, 068101 (2007).

[17] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **107**, 21441 (2010).

[18] A. I. Zayed, Handbook of Function and Generalized Function Transformations (CRC Press, Boca Raton, LA, 1996).

[19] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, Proc. Natl. Acad. Sci. U.S.A. **103**, 6190 (2006).

[20] D. P. Giedroc and P. V. Cornish, Virus. Res. **139**, 193 (2009).

[21] I. Brierley, R. J. C. Gilbert, and S. Pennell, in Recoding: Expansion of Decoding Rules Enriches Gene Expression, edited by J. F. Atkins and R. F. Gesteland, Nucleic Acids and Molecular Boology (Springer, New York, 2010), Vol. 24, pp. 149–174.

[22] T. R. Cech, Cell **116**, 273 (2004).

[23] C. A. Theimer, L. D. Finger, L. Trantirek, and J. Feigon, Proc. Natl. Acad. Sci. U.S.A. **100**, 449 (2003).

[24] E. D. Holmstrom and D. J. Nesbitt, J. Phys. Chem. B **118**, 3853 (2014).

[25] X. Qu, J.-D. Wen, L. Lancaster, H. F. Noller, C. Bustamante, and I. Tinoco, Nature (London) **475**, 118 (2011).

[26] D. B. Ritchie, D. A. N. Foster, and M. T. Woodside, Proc. Natl. Acad. Sci. U.S.A. **109**, 16167 (2012).

[27] D. B. Ritchie, J. Soong, W. K. A. Sikkema, and M. T. Woodside, J. Am. Chem. Soc. **136**, 2196 (2014).

[28] L. Green, C.-H. Kim, C. Bustamante, and I. Tinoco, J. Mol. Biol. **375**, 511 (2008).

[29] G. Chen, J.-D. Wen, and I. Tinoco, Jr., RNA **13**, 2175 (2007).

[30] K. H. White, M. Orzechowski, D. Fourmy, and K. Visscher, J. Am. Chem. Soc. **133**, 9775 (2011).

[31] M. de Messieres, J.-C. Chang, A. T. Belew, A. Meskauskas, J. D. Dinman, and A. La Porta, Biophys. J. **106**, 244 (2014).

[32] P. V. Cornish, M. Hennig, and D. P. Giedroc, Proc. Natl. Acad. Sci. U.S.A. **102**, 12694 (2005).

[33] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, Biophys. J. **72**, 1335 (1997).

[34] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).

[35] M. J. Lang, P. M. Fordyce, A. M. Engh, K. C. Neuman, and S. M. Block, Nat. Methods **1**, 133 (2004).

[36] P. T. X. Li, C. Bustamante, and I. Tinoco, Jr., Proc. Natl. Acad. Sci. U.S.A. **103**, 15847 (2006).

[37] K. Neupane, H. Yu, D. A. N. Foster, F. Wang, and M. T. Woodside, Nucleic Acids Res. **39**, 7677 (2011).

[38] R. Zwanzig, Proc. Natl. Acad. Sci. U.S.A. **85**, 2029 (1988).

[39] A. Solanki, K. Neupane, and M. T. Woodside, Phys. Rev. Lett. **112**, 158103 (2014).

[40] P. P. Chapagain, J. L. Parra, B. S. Gerstman, and Y. Liu, J. Chem. Phys. **127**, 075103 (2007).

[41] K. Neupane, A. Solanki, I. Sosova, M. Belov, and M. T. Woodside, PLoS One **9**, e86495 (2014).

[42] W. J. Greenleaf, M. T. Woodside, and S. M. Block, Annu. Rev. Biophys. Biomol. Struct. **36**, 171 (2007).

[43] J. Dong, C. E. Castro, M. C. Boyce, M. J. Lang, and S. Lindquist, Nat. Struct. Mol. Biol. **17**, 1422 (2010).