Measuring aptamer folding energy using a molecular clamp

Hao Qu^{a,*}, Qihui Ma^a, Lu Wang^a, Yu Mao^a, Michael Eisenstein^{b,c},

Hyongsok Tom Soh^{b,c,*}, Lei Zheng^a

^a School of Food and Biological Engineering, Hefei University of Technology, Hefei, 230009, China

^b Department of Electrical Engineering, Stanford University, Stanford, CA 94305

^c Department of Radiology, Stanford University, Stanford, CA 94305

* Corresponding authors at:

Hefei University of Technology, Hefei, 230009, Anhui, P. R. China. Tel.: +86 551 62901516, and Stanford University, Stanford, CA 94305

E-mail addresses: quhao@hfut.edu.cn (H. Qu); tsoh@stanford.edu (H. T. Soh).

Abstract

Folding energy offers a useful metric for characterizing the stability and function of aptamers. However, experimentally measuring the folding energy is challenging and there is currently no general technique to measure this parameter directly. In this work, we present a simple approach for measuring aptamer folding energy. First, the aptamer is stretched under equilibrium conditions with a double-stranded DNA “molecular clamp” that is coupled to the aptamer ends. We then measure the total internal energy of stressed DNA molecules using time-lapse gel electrophoresis and compare the folding and unfolding behavior of molecular clamp-stressed molecules that incorporate either the aptamer or unstructured random single-stranded DNA in order to derive the aptamer folding energy. Using this approach, we measured a folding energy of 10.40 kJ/mol for the HD22 thrombin aptamer, which is consistent with other predictions and estimates. We also analyzed a simple hairpin structure, generating a folding energy result of 9.05 kJ/mol, consistent with the value predicted by computational models (9.24 kJ/mol). We believe our strategy offers an accessible and generalizable approach for obtaining such measurements with virtually any aptamer.

Keywords: Folding energy; Equilibrium; Aptamer; Molecular Clamp

Introduction

Aptamers offer the powerful capability to recognize and specifically bind to a wide range of biomolecules, including small molecules ¹, metal ions ², proteins ³, and even whole cells ⁴. An aptamer's function is dependent upon its capacity to fold into and maintain a binding-competent conformation ⁵. As such, the folding energy for a given DNA or RNA sequence is an important determinant of aptamer stability, which in turn affects that aptamer's target affinity and specificity, and thus its broader utility in molecular detection and other assays.

Unfortunately, it currently remains difficult to evaluate this parameter because of the difficulties associated with direct measurement of aptamer folding energy at equilibrium. One approach is based on time-dependent single-molecule pulling experiments, in which oligonucleotides are stretched by atomic force microscopy (AFM) or optical tweezers ⁶⁻⁹. These experiments drive the system away from equilibrium, where the magnitude of the rupture force is related to the pulling rate. As a consequence, one must use the relationship between free energies and irreversible work ¹⁰⁻¹² described by Jarzynski's equality ¹³ to calculate free energy from repeated measurements. However, some have challenged the validity of this approach, arguing that its underlying assumption—*i.e.*, the connection between the microscopic work performed by a time-dependent force with the corresponding Hamiltonian—is theoretically inconsistent ¹⁴⁻¹⁶. Other groups have used ultraviolet spectroscopy or circular dichroism (CD) to generate indirect measurements of melting curves based on distinct differences in the spectroscopic properties of folded and unfolded aptamers ^{17,18}.

However, the arbitrary determination of melting curve baselines can introduce considerable uncertainty, and the assumption that denaturation is only a two-state process with zero heat-capacity change greatly limits the universal application of this approach^{19,20}. Others have also performed melting experiments using differential scanning calorimetry (DSC),^{18,20} which measures differential heat capacity as a function of temperature^{19,20}. However, the accuracy of this approach is not guaranteed, because the ability to correctly determine the values of many thermodynamic parameters is heavily dependent on the choice of appropriate melting models^{20,21}.

We describe here a general strategy for directly measuring aptamer folding energy, which eliminates these confounders and sources of bias that can lead to inaccurate measurement. Our method employs a short piece of double-stranded (ds) DNA, which is shorter than the dsDNA persistence length of ~50 nm, as a “molecular clamp” to pull an aptamer into an unfolded state under equilibrium conditions. Once the appropriate length of the molecular clamp required for successful extension of the aptamer has been determined, we measure the total energy of the constrained molecule – including the bending energy of the molecular clamp and stretching energy of the aptamer – through a simple time-lapse gel electrophoresis method. The folding energy is calculated by comparing against the gel profile of a randomized, unstructured single-stranded (ss) DNA molecule that has been incubated with the same molecular clamp. Here, we demonstrate this method to determine the folding energy of the thrombin aptamer HD22²², and derive a measurement that is consistent with prior estimates. Our ‘folding energy measurement by molecular clamping’ (FE-MC) method only requires standard gel

electrophoresis equipment and reagents, and therefore offers a simple and broadly accessible solution for characterizing aptamer folding and stability.

Results and Discussion

Overview of the FE-MC method

The central idea behind the FE-MC method is to exploit short stretches of dsDNA that act as “molecular clamps” to unfold aptamers. This requires the design of two DNA oligonucleotides: Strand A contains the aptamer sequence at the center, and is flanked by sequences that are complementary to Strand B (**Fig. 1A**, left). Therefore, hybridization between Strand A and Strand B produces a circularized, partially double-stranded molecule (**Fig. 1A**, right). This structure can be viewed as two coupled (nonlinear) springs – representing the dsDNA and ssDNA component – which are constrained at the same end-to-end distance (EED) x . This represents the distance between the two ends of the molecule after either bending for the dsDNA or stretching for the ssDNA, and describes the degree of deformation for both components. We have termed this circularized molecule a ‘stressed aptamer molecule’ (SAM), with a high internal energy that is stored in the bending energy of the dsDNA component (E_d) and the stretching energy of the ssDNA component (E_s). The dsDNA component applies stretching force at pico-Newton (pN) scales on the ssDNA component, *i.e.* the aptamer sequence, at equilibrium ²³, the stretching behavior of which is modeled as shown in **Fig. S1A**. Since shorter molecular clamps would be expected to exert a stronger extending force (as explained in the Supplementary Information, SI), the aptamer may

be successfully stretched into an unfolded state (**Fig. 1A**, lower right) or remain folded (**Fig. 1A**, upper right) depending on which state has lower internal energy for varying length of the dsDNA component (N_d), as illustrated in **Fig. S1B**.

The next step entails identification of the appropriate N_d value to ensure that the aptamer is fully extended upon hybridization. To achieve this, we construct a series of SAMs with molecular clamps of varying N_d and subject these constructs to gel electrophoresis. The resulting gel will show multiple bands corresponding to monomers, dimers, and other multimers arising from polymerization of multiple copies of Strand A and Strand B. In this step, we need to identify the value of N_d at which splitting of the monomer band is observed (**Fig. 1B**), indicating the coexistence of two possible configurations with similar internal energy (as depicted in **Fig. S2**). This represents the transition length (N_c) of the molecular clamp: when $N_d < N_c$, the aptamer is stretched and unfolded, whereas when $N_d > N_c$, the aptamer remains in a folded state.

Figure 1. Determination of aptamer folding energy using the molecular clamp (FE-MC) method. (A) This procedure entails constructing a series of stressed aptamer molecules (SAMs). These comprise two strands: Strand A contains an aptamer sequence at its center (yellow), flanked by sequences (cyan) complementary to Strand B (red). Hybridization produces a circularized structure with a ‘break’ in one strand in the middle of the dsDNA component. The dsDNA component (with length N_d) serves as an ‘molecular clamp’ that applies pN forces on the ssDNA aptamer component (with length N_s). There is high internal energy stored in the dsDNA bending energy and ssDNA stretching energy, and depending on the length of N_d , the aptamer may be stretched into an unfolded state (upper right) or remain folded (lower right). (B) The next step of FE-MC is to find the critical molecular clamp length (N_c) at which the aptamer transitions from folded to unfolded. This entails gel electrophoresis of SAMs with different N_d values, yielding bands corresponding to monomers, dimers and other multimers arising from polymerization of multiple copies of Strand A and Strand B. N_c can be identified as the sample in which splitting of the monomer band is observed (red dashed box), indicating the coexistence of folded and unfolded aptamer states. A SAM in which $N_d < N_c$ is chosen for subsequent steps. (C) We next measure the internal energy (ΔG_{apt}) of the chosen SAM. Because of the break in the molecular clamp, this internal energy may be relaxed by forming dimers, in which the dsDNA component does not bend and the aptamer is not stretched. We can then quantify the SAM internal energy using Eq. (1) by measuring the equilibrium concentrations of monomers and dimers. Due to the concentration bias induced by ongoing monomer-dimer interconversion even during the gel electrophoresis, we obtain initial equilibrium concentrations by running time-lapse gel electrophoresis with successively delayed loading of multiple lanes with aliquots from the same SAM sample, and fitting ‘snapshots’ of gel profiles (red curves) with the reaction-diffusion model (blue curves). (D) In parallel, we construct stressed control molecules (SCMs) in which the aptamer from the SAM is substituted with an unstructured randomized sequence, and again measure the internal energy (ΔG_{ran}). (E) This plot depicts the energy levels occupied by the monomeric and dimeric states of the SAM and SCM. Note that the energy level is the same for the unfolded SAM and SCM monomers, but the dimer energy levels are different because the aptamer is folded while the randomized sequence is unstructured. The aptamer folding energy ΔG_{fold} is calculated as $\Delta G_{apt} - \Delta G_{ran}$.

We then select a SAM for which the molecular clamp length N_d is smaller than N_c for further analysis. It is important to note that there is a break in the middle of the dsDNA component of the SAM, which arises from the two ends of the aptamer-flanking ‘arms’ of Strand A after hybridization with Strand B. This break means that the SAM may relax its internal elastic energy through the formation of multimers, in which the dsDNA component is not forced to bend and the ssDNA component is not subject to

stretching²⁴. Because polymerization lowers the entropy of molecules, there exists a chemical equilibrium of monomer-dimer interconversion, which can be observed through gel electrophoresis, with multiple bands corresponding to dimers and other multimers (**Fig. 1C**).

Quantitatively, this monomer-dimer interconversion can be described in terms of the equality of monomer and dimer chemical potentials, namely $2\mu_M = 2\Delta G_{M-\frac{1}{2}D} + 2k_B T \ln X_M$ and $\mu_D = k_B T \ln X_D$. The internal energy of the SAM is hence described by the following equation:

$$\Delta G_{M-\frac{1}{2}D} = \frac{1}{2} k_B T \ln \left(\frac{X_D}{X_M^2} \right) = \frac{1}{2} k_B T \ln \left[\frac{C_D/C_W}{(C_M/C_W)^2} \right] \quad (1)$$

where X_D , and X_M are the molar fractions and C_D and C_M are the concentrations of the monomer and dimer, respectively. $C_W = 55$ M is the concentration of water (see Methods for deriving details). Eq. (1) suggests that the internal energy of the SAM can simply be computed by quantification of monomer and dimer populations at equilibrium (*e.g.*, through gel electrophoresis). It is worth noting that this monomer-dimer interconversion continues even during gel electrophoresis, leading to bias in monomer and dimer concentrations relative to initial equilibrium values. To address this problem, we have adopted a ‘time-lapse’ gel electrophoresis method (**Fig. 1C**, right). Specifically, multiple aliquots of the same chosen SAM sample (in which $N_d < N_c$) are loaded into different lanes at successive times. The gel image for this SAM is then collected to show the evolution of monomer and dimer bands with time, enabling extrapolation back to initial equilibrium concentrations. A reaction-diffusion model is applied with the initial monomer and dimer concentrations as fitting parameters, and

their values are determined so that the gel profiles at each running time are fitted to model curves with the same set of parameters (**Fig. 1C**, right; see Methods for details). The initial concentrations are then plugged into Eq. (1) to calculate the internal energy ΔG_{apt} within the SAM.

Next, we construct stressed control molecules (SCMs), in which the aptamer sequence in Strand A has been replaced with a randomized sequence (**Fig. 1D**). We extract the internal energy within the SCM (ΔG_{ran}) in the same way as for the SAM. Finally, we compute the aptamer folding energy as follows: $\Delta G_{fold} = \Delta G_{apt} - \Delta G_{ran}$ (**Fig. 1E**). Since the aptamers in the SAM dimers are in a folded state while the randomized sequences in the SCM dimers are in a random-coiled state, $\Delta G_{apt} - \Delta G_{ran}$ cancels the contribution of base-pairing in half of the SAM and SCM dimers, and yields the energy difference between the folded aptamer and the random DNA coil (see SI).

Folding energy measurement of the HD22 thrombin aptamer

To demonstrate the FE-MC method, we chose an aptamer that recognizes human α -thrombin (HD22)²². This 29-nt aptamer (5'-AGTCCGTGGTAGGGCAGGTTGGGGTGACT-3') contains a duplex/G-quadruplex mixed structure²⁵, and exhibits strong target affinity ($K_D \sim 0.5$ nM). We constructed a series of HD22 SAMs with N_d ranging from 18–24 bp. Time-lapse gel images of these SAMs are shown in **Fig. 2**. When $N_d = 20$ bp, splitting of the monomer bands was observed (**Fig. 2B**), indicating that this was the N_c of the molecular clamp; when $N_d < 20$, the HD22 aptamer was unfolded, whereas when $N_d > 20$, the HD22 aptamer

remained in the folded state.

We selected an unfolded SAM ($N_d = 18$ bp; SAM18) for subsequent detailed analysis. A time-lapse gel image of SAM18 was collected as shown in **Fig. 2A** and the gel profiles in different lanes (*i.e.*, at different running times) were successfully fitted with the reaction-diffusion model (**Fig. S3A**, see SI for fitting details). We found that the initial equilibrium concentrations of monomer and dimer were 0.42 μM and 0.79 μM , respectively. We hence calculated the internal energy of the SAM18 monomer $\Delta G_{SAM18} = \frac{1}{2} k_B T \ln \left[\frac{0.79 \mu\text{M}/55 \text{ M}}{(0.42 \mu\text{M}/55 \text{ M})^2} \right] = 9.96 k_B T = G_{Monomer}^{SAM18} - \frac{1}{2} G_{Dimer}^{SAM18}$ using Eq. (1). This describes the free energy difference between the SAM18 monomer ($G_{Monomer}^{SAM18}$) and half of the SAM18 dimer ($\frac{1}{2} G_{Dimer}^{SAM18}$).

Figure 2. Time-lapse gel images for SAMs with the HD22 aptamer and N_d = A) 18, B) 20, C) 22, and D) 24 bp. (E) Time-lapse gel pictures for an SCM with a randomized control sequence and N_d = 18.

We next constructed SCM18, with the same molecular clamp, but where Strand A incorporated a control 29-mer randomized sequence (5'-CAGCAGGCAATCGATACACACACAGTAGA-3') with no secondary structure (checked by the DINAMelt web server ²⁶). The time-lapse gel image was collected as shown in **Fig. 2E**. The internal molecular energy was calculated using Eq. (1) by fitting the SCM18 gel profiles with the reaction-diffusion model (**Fig. S3B**). We found that

the initial monomer and dimer concentrations for the SCM were 1.78 μM and 0.11 μM , respectively (see SI for fitting details). Thus, $\Delta G_{SCM18} = 5.62 k_B T = G_{Monomer}^{SCM18} - \frac{1}{2} G_{Dimer}^{SCM18}$.

SAM18 and SCM18 shared the same free energy in their stretched state, but the aptamers in the SAM18 dimers were folded while the randomized sequences in the SCM18 dimers were in a random-coiled state, leading to a difference in energy levels. Accordingly, the differential between ΔG_{SAM18} and ΔG_{SCM18} canceled the base-pairing energy in half of the SAM and SCM dimers, yielding the energy difference between the folded aptamer and the random DNA coil. Specifically, this can be calculated as follows:

$$\Delta G_{SAM18} - \Delta G_{SCM18} = \frac{1}{2} G_{Dimer}^{SCM18} - \frac{1}{2} G_{Dimer}^{SAM18} = \Delta G_{fold}^{HD22} = 4.34 k_B T$$

or 10.40 kJ/mol for the HD22 folding energy.

Validation of the experimental results

In order to validate this measurement, we calculated the critical molecular clamp length (N_c) to check whether it was consistent with experimental observations. The theoretical value of N_c was determined for a scenario in which the folded and extended states of the SAM have similar energy values. Our determination of the folding energy for HD22 (4.34 $k_B T$) was plugged into the model of the aptamer folding energy profile (see SI) as the ssDNA stretching energy of the SAM. The EED x between the two ends of the SAM was used to characterize the degree of deformation for the two SAM components. The total energy (E_{tot}) of the SAM as a function of the EED x was

calculated using the sum of the ssDNA stretching energy and dsDNA bending energy (see SI). We found that when $N_d = 20$, the local energy minimum of the folded state was 4.95 k_BT, while that of the stretched state was 5.20 k_BT, as shown in **Fig. S2**. The similarity of the two energy values indicates that $N_d = 20$ is the theoretical critical length, consistent with our experimental results.

We also compared the final HD22 folding energy value 4.34 k_BT (or 10.40 kJ/mol) to other estimates. Prior studies of aptamer folding based on pulling experiments using AFM or optical tweezers have yielded free energy estimates for the complete unfolding of the *pbuE* riboswitch aptamer in the range of 2–3 kcal/mol (8.4–12.6 kJ/mol)⁹, the same order of magnitude for energy scale as we obtained with HD22 using our orthogonal method. For this particular aptamer, the DINAMelt web server²⁶ predicts a folding energy of just 1.11 kcal/mol (4.67 kJ/mol), which is much smaller than the value we obtained. This is because the DINAMelt algorithm does not account for the energy of the 3D structure of the G-quadruplex. According to previous studies using UV-spec or DSC, the ultimate free energy $\Delta G(310\text{ K})$ for a low-order G-quadruplex is approximately 1.50 kcal/mol (6.30 kJ/mol)^{17,19}. Since HD22 features a combination of duplex and G-quadruplex structure²⁵, a rough estimate from adding the base-pairing energy of the duplex (4.67 kJ/mol) and the G-quadruplex folding energy (6.30 kJ/mol) yields a total folding energy of 10.97 kJ/mol for HD22, which is very similar to the value we calculated (10.40 kJ/mol).

Such an estimate is necessary given the lack of directly comparable, precise folding energy predictions by other metrics, but we were also able to confirm the accuracy of

our FE-MC approach with a simple hairpin structure (5'-TTGTCAT TTTTTTTTTTTTTTTATGACTT-3'). The folding energy of this simple structure is easily predicted by DINAMelt ²⁶ to be 9.24 kJ/mol under ambient conditions, and FE-MC yielded a highly consistent measurement of 9.05 kJ/mol (see SI for details), further supporting the validity and accuracy of our approach.

4. Conclusion

In this work, we describe a simple approach for probing the folding energy of aptamers. In our ME-FC method, the aptamer is unfolded by a dsDNA component that acts as a molecular clamp under equilibrium conditions. The internal energy of this stressed aptamer molecule is determined using time-lapse gel electrophoresis, and then compared against a similar stressed DNA molecule in which the aptamer has been substituted with a randomized ssDNA with no secondary structure. The internal energy difference between the two stressed molecules yields the aptamer folding energy.

Using this approach with the HD22 thrombin aptamer as a model, we computed a folding energy of 4.34 k_BT (~10.4 kJ/mol), which is in keeping with aptamer folding energy scales from other studies as well as rough predictions based on HD22's structure. We also validated the accuracy of FE-MC by correctly measuring the folding energy of a hairpin structure, obtaining results that mirror the value predicted by the DINAMelt web server. Although we have calculated the folding energy measurement of a relatively short aptamer (29 nt) here, FE-MC method should also be applicable to aptamers with longer lengths of 60–80 nt. This is valuable in that many newly-selected aptamers

include primer-binding sites and have not been minimized. Since FE-MC only requires time-lapse gel electrophoresis experiments, we believe it offers a highly accessible and generalizable method for evaluating aptamers in a simple and direct manner.

Methods and Materials

Sample preparation

All oligonucleotide sequences (listed in **Table 1**) were synthesized by Sangon Biotech (Shanghai) with HPLC purification. The red sequences of Strand A hybridize with the corresponding Strand B. We confirmed hybridization of these dsDNA regions by checking sequence thermodynamic properties with the DINAMelt web server^{26,27}.

Table 1. DNA sequences used to form SAMs with different N_d .

	<i>Strand A (5' to 3')</i>	<i>Strand B (5' to 3')</i>
Group 1 ($N_d = 24$)	ACG TGA GAG CAG AGTCCGTGGTAGGGCAGGTTGGGGTG ACT GAC ATA CGA CGA	CTG CTC TCA CGT TCG TCG TAT GTC
Group 2 ($N_d = 22$)	ACG TGA GAG CA AGTCCGTGGTAGGGCAGGTTGGGGTG ACT AC ATA CGA CGA	TG CTC TCA CGT TCG TCG TAT GT
Group 3 ($N_d = 20$)	ACG TGA GAG C AGTCCGTGGTAGGGCAGGTTGGGGTG ACT C ATA CGA CGA	G CTC TCA CGT TCG TCG TAT G
Group 4 ($N_d = 18$)	ACG TGA GAG AGTCCGTGGTAGGGCAGGTTGGGGTG ACT ATA CGA CGA	CTC TCA CGT TCG TCG TAT

Each Strand A was mixed with equimolar amounts of Strand B and diluted to a final concentration of 2 μ M and a final volume of 200 μ L in the hybridizing buffer (10 mM Tris, 5 mM MgCl₂, 100 mM NaCl, pH = 7.9). The mixture was then heated at 95 °C for 10 min and annealed at room temperature overnight (~0.065 °C/min) to ensure the proper conformational structures in stressed DNA molecules.

Molecular energy extraction based on monomer-dimer equilibrium

At equilibrium of monomer-dimer interconversion, the chemical potentials for two

monomers ($2\mu_M = 2\Delta G_{M-\frac{1}{2}D} + 2k_B T \ln X_M$) and one dimer ($\mu_D = k_B T \ln X_D$) are equal: $2\Delta G_{M-\frac{1}{2}D} + 2k_B T \ln X_M = k_B T \ln X_D$. Since the base-pairing is identical for one dimer and two monomers, the energetic contribution of base-pairing cancels out and the internal elastic energy stored in the monomer can be characterized by the following equation: ²⁴

$$\Delta G_{M-\frac{1}{2}D} = \frac{1}{2} k_B T \ln \left(\frac{X_D}{X_M^2} \right)$$

where $X_D = C_D/C_W$ and $X_M = C_M/C_W$, describing the molar fraction of the monomer and dimer respectively, and C_M and C_D are the respective concentration of the two species. $C_W = 55$ M is the concentration of water. This equation essentially gives the energy difference between the monomer and half of the dimer (as the ground state), and suggests that this energy difference can be simply extracted by measuring the equilibrium monomer and dimer concentrations.

Time-lapse gel electrophoresis

Concentrations of monomers and dimers were experimentally determined by time-lapse gel electrophoresis with a 5% polyacrylamide gel, based on the intensities of the corresponding gel bands (using necessary calibrations as described in ²⁴). The monomer-dimer interconversion of the stressed DNA molecules also occurs while the sample is moving through the gel and has to be taken into account. Therefore, in order to extrapolate back the initial equilibrium concentrations of the population of monomers and dimers at t_0 , we loaded different lanes of the gel with aliquots from the same sample at intervals of 10 min to capture “snapshot” gel profiles at different times.

These various aliquots were run under 100 V for 55 min, 45 min, 35 min, 25 min, and 15 min, respectively. Images from time-lapse gel electrophoresis were converted to gel profiles using ImageJ ver1.51.

Reaction-diffusion model

In order to extract the initial population of monomers and dimers, we fitted these gel profiles with a reaction-diffusion model²⁸ that has been successfully applied to the analysis of electrophoresis with protein-DNA²⁹, hairpin-duplex³⁰, and monomer-dimer^{23,24} interconversions. The model formulas are:

$$\begin{cases} \frac{\partial C_m}{\partial t} = D_m \frac{\partial^2 C_m}{\partial x^2} - V_m \frac{\partial C_m}{\partial x} - 2k_1 C_m^2 + 2k_2 C_d \\ \frac{\partial C_d}{\partial t} = D_d \frac{\partial^2 C_d}{\partial x^2} - V_d \frac{\partial C_d}{\partial x} + k_1 C_m^2 - k_2 C_d \end{cases}$$

where C_m and C_d are the concentration of monomers and dimers respectively, as a function of gel running time t and gel vertical position x from the bottom of the loading well; D_m and D_d are the diffusion constant of the two species, and V_m and V_d are their mobilities; k_1 is the conversion rate from monomers to dimers, while k_2 is the rate for the reverse reaction. During fitting with the reaction-diffusion model, V_m and V_d were first determined by measuring the distance that the population had shifted at various running times. D_m and D_d were determined from the width of the corresponding peaks of monomers and dimers. The initial monomer and dimer concentrations (C_{m0} and C_{d0}) and interconversion rates (k_1 and k_2) were adjusted so that the model could fit the gel profiles at different times (*i.e.*, in different lanes) with the same set of parameters. Finally, C_{m0} and C_{d0} were used to calculate the internal energy of the stressed DNA molecules.

Note that the computed elastic energy was rather insensitive to the parameter values in the model. Even a barely fair fitting to the gel profiles, in which the interconversion rates were set to zero, leads to a difference of $\sim 15\%$ in the initial concentration values, which in turn resulted in only a $\sim 0.2 k_B T$ change in the final elastic energy result due to the logarithmic function in Eq. (1).

Acknowledgements

This study was supported by the National Key R&D Program of China (2017YFC1600603), the National Natural Science Foundation of China (21705031), the Natural Science Foundation of Anhui Province (1808085QB39), the Fundamental Research Funds for the Central Universities (PA2019GDQT0018), and Engineering Research Center of Bio-process by Ministry of Education in Hefei University of Technology. This work was also supported by the Chan-Zuckerberg Biohub (HTS).

Supplementary Information

Hao Qu^{a,*}, Qihui Ma^a, Lu Wang^a, Yu Mao^a, Michael Eisenstein^{b,c},

Hyongsok Tom Soh^{b,c,*}, Lei Zheng^a

^a School of Food and Biological Engineering, Hefei University of Technology, Hefei, 230009, China

^b Department of Electrical Engineering, Stanford University, Stanford, CA 94305

^c Department of Radiology, Stanford University, Stanford, CA 94305

* Corresponding authors at:

Hefei University of Technology, Hefei, 230009, Anhui, P. R. China. Tel.: +86 551 62901516, and Stanford University, Stanford, CA 94305

E-mail addresses: quhao@hfut.edu.com (H. Qu); tsoh@stanford.edu (H. T. Soh).

Use of dsDNA as molecular clamp

Long chains of dsDNA normally exhibit bending elasticity resembling an entropic spring, in accordance with the ‘wormlike chain’ model (WLC) with energy of:

$$E_{WLC} = \int_0^{2L} \frac{1}{2} \frac{B}{R(s)} ds$$

where R is the radius of curvature, s the arc length along the rod, B the bending modulus ($B \approx 200 \text{ pN} \times \text{nm}^2$), and $2L$ the contour length ($N_d \times 0.33 \text{ nm}$)^{31,32}. However, this model breaks down when the contour length is less than the persistence length of dsDNA (*i.e.*, 150 base pairs, or 50 nm). We have previously determined that this breakdown involves a local “kink” forming in the middle of the dsDNA, which may greatly lower the bending energy of the molecule after a critical deformation. The following is an analytical expression of the dsDNA bending energy (E_d) at this length:

$$E_d(x) = \begin{cases} \tau_c \arccos\left(\frac{x}{2R}\right) & \text{for } 0 < x < x_c \\ \frac{5B}{L} \frac{x_0 - x}{2L} - k_B T \ln\left(\frac{2L - x}{2L - x_0}\right), & \text{for } x_c < x < x_0 \end{cases} \quad (\text{S1})$$

where τ_c is the critical torque at which the kink develops; x the end-to-end distance (EED); $R = L(1 - 2\gamma^2/45)$ is the arc length, where $\gamma = L\tau_c/(2B)$; $x_0 = 2L[1 - k_B TL/(5B)]$ is the EED at zero force; and x_c is the critical EED^{23,24,33}. The lower form of the equation corresponds to the smoothly bent solution, while the upper describes the kinked state. Therefore, the bending behavior – including the applicable force for using short dsDNA as molecular clamp – is comprehensively known. For example, according to Eq. (S1), the bending elastic energy of dsDNA shows an increase with decreasing length (N_d). Hence, the stretching state of the aptamer can be readily tuned by adjusting N_d .

Model of aptamer folding energy profile

The stretching force applied to the aptamer in the context of the clamp initially leads to disruption and unfolding of its structure, with a corresponding free energy ΔG_{fold} . After the aptamer is completely unfolded at a critical deformation, it returns to a coiled state as simple ssDNA without any internal structures, and the free energy required for pulling is purely elastic, arising from the entropy change of the ssDNA as a random coil (ΔG_{coil}). The random coil state follows the WLC model described by the polynomial expansion of the Marko-Siggia expression ⁷:

$$\Delta G_{coil} = \frac{9k_B T}{4N_s l_s^2} \left(x^2 + \frac{x^3}{N_s l_s} + \frac{3x^4}{N_s^2 l_s^2} \right) \quad (S2)$$

where x is the EED, N_s the number of bases of ssDNA, and l_s the persistence length of ssDNA (~ 0.7 nm). Therefore, the stretching energy for the aptamer can be modeled as the superposition of an energy well with a depth of ΔG_{fold} in the elastic energy of random coil ΔG_{coil} , as described in Eq. (S2) ³¹ and shown in **Fig. S1A**. The detailed energy landscape of aptamer folding energy is not discussed in this paper.

The effect of molecular clamps on aptamer folding states

For FE-MC to succeed, the dsDNA molecular clamp must be capable of pulling the aptamer apart from the folded state. To understand the effect of molecular clamps on aptamer folding states, we have calculated the total energy profile using the energy model for stretching energy of the ssDNA component (**Fig. S1A**), with the depth of the energy well of the experimentally measured HD22 aptamer folding energy (4.34 k_BT), and Eq. (S1) for bending energy of the dsDNA component. We found that aptamers

situated within stressed DNA molecules may shift from a folded to an extended state as the length of the molecular clamp decreases, producing “stiffer” clamps. As simulated in **Fig. S1B**, there are two local minimums for the total energy vs EED for stressed DNA molecules containing the aptamer, corresponding to the stretched and folded states respectively. When $N_d = 22$, the local minimum with a small separation (*i.e.*, EED = 1.02 nm) is lower, and the energy partition is $E_d = 9.39$ k_BT in the molecular clamp and $E_s = -4.19$ k_BT in the aptamer, where the majority of the energy lies in the molecular clamp. But when $N_d = 18$, the local minimum with a large separation (*i.e.* EED = 5.92 nm) becomes lower, and the energy partition is $E_d = 0.01$ k_BT in the molecular clamp and $E_s = 4.76$ k_BT in the aptamer. In this context, the majority of the energy is stored in the aptamer, leading to its full extension.

Figure S1. (A) The energy model for aptamer stretching: the superposition of an energy well with depth of ΔG_{fold} on the WLC energy ΔG_{coil} as expressed by Eq. (S2). (B) Calculated total energy profile (using Eq. (S1) and the aptamer energy model in A) vs EED for stressed DNA molecules with the HD22 aptamer and a molecular clamp of $N_d = 22$ (blue) or 18 (orange). The global energy minimum shifts from the local minimum on the left to the one on the right as N_d changes from 22 to 18, corresponding to the transition from the compressed state to the stretched state.

Comparison of the experimental results with a theoretical model

For validation, the folding energy value obtained for the HD22 aptamer with FE-MC (4.34 k_BT) was plugged into the theoretical model for energy profiles of SAMs to check whether it would produce the same critical molecular clamp length (N_c) as we observed in the time-lapse gel. The total energy of the SAM with the HD22 aptamer as a function of the EED x (E_{tot} vs x) was calculated by taking the sum of the dsDNA bending energy (Eq. (S1)) and ssDNA stretching energy (see aptamer folding energy profile model above). The total energy profile for the SAM with $N_d = 20$ is shown in **Fig. S2**. The two local energy minimums correspond to the folded and stretched states, with energy values of 4.95 k_BT and 5.20 k_BT, respectively. This clearly indicates that the critical molecular clamp length $N_c = 20$, is consistent with our measurements after varying N_d as shown in **Fig. 2**, confirming the validity of the FE-SM method.

Figure S2. Calculated total energy profile (using Eq. (S1) and the aptamer energy model) vs EED for the SAM with the HD22 aptamer and a molecular clamp of $N_d = 20$. The measured HD22 folding energy value 4.34 k_BT was plugged into the aptamer energy model for energy well depth. The two local energy minimums corresponding to the folded and stretched states were calculated to be 4.95 k_BT and 5.20 k_BT, respectively. The similarity of these two values indicates that the critical length $N_c = 20$, consistent with the experimental results shown in **Fig. 2**.

Fitting details using the reaction-diffusion model

As described in Methods, the reaction-diffusion model was used for analysis of the time-lapse gel profiles for SAM18 and SCM18. For the SAM18 gel profiles, the vertical positions of the monomer and dimer bands (from the bottom of the loading well) at different running time were determined so that the mobilities for both species could be extracted using a linear fitting. We calculated $V_m = 10.26 \times 10^{-2}$ cm/min and $V_d = 7.59 \times 10^{-2}$ cm/min, respectively. The diffusion constants D_m and D_d were set according to the width of the corresponding peaks in the gel profiles and found to be $D_m = 2.02 \times 10^{-5}$ cm²/min and $D_d = 1.95 \times 10^{-5}$ cm²/min. The other parameters were adjusted so that the model might fit the experimental gel profiles at different times with identical values for these parameters. During the actual fitting, the heights of the monomer and dimer peaks were the key fitting factors. It should be noted that multiple sets of parameters may produce comparable fitting results for the gel profile at a given running time, but only a unique combination of parameters would work for all running times. For the gel profile of SAM18, $C_{m0} = 0.42$ μ M, $C_{d0} = 0.79$ μ M, $k_1 = 0.0001$ μ M⁻¹ min⁻¹ (monomer association rate), and $k_2 = 0.0249$ min⁻¹ (dimer dissociation rate) generated good fitting results for the gel profiles at 55, 45, and 35 min, as shown in **Fig. S3A**.

Gel profile analysis for SCM18 was slightly more complicated because there were two monomer bands in the gel patterns. Therefore, besides the interconversion between dimers and monomers, the interconversion between two monomer states had also to be considered. Following the same fitting procedures, the mobilities for the two monomer species and dimers were obtained depending on the vertical positions of the

corresponding bands ($V_{m1} = 6.87 \times 10^{-2} \text{ cm/min}$, $V_{m2} = 6.45 \times 10^{-2} \frac{\text{cm}}{\text{min}}$, and $V_d = 5.43 \times 10^{-2} \text{ cm/min}$). The diffusion constants were identified to be $D_{m1} = 1.95 \times 10^{-5} \text{ cm}^2/\text{min}$, $D_{m2} = 1.95 \times 10^{-5} \text{ cm}^2/\text{min}$, and $D_d = 1.37 \times 10^{-5} \text{ cm}^2/\text{min}$. By adjusting the other fitting parameters, we found that the combination of $C_{m10} = 1.63 \text{ } \mu\text{M}$, $C_{m20} = 0.15 \text{ } \mu\text{M}$, $C_{d0} = 0.11 \text{ } \mu\text{M}$, $k_1 = 0.0040 \text{ } \mu\text{M}^{-1} \text{ min}^{-1}$ (monomer association rate), $k_2 = 0.0027 \text{ min}^{-1}$ (dimer dissociation rate), and $k_3 = 0.0100 \text{ min}^{-1}$ (monomer interconversion rate) would produce the best fitting results for the gel profiles at 55, 45, and 35 min, as shown in **Fig. S3B**.

Figure S3. Intensity profiles of the 55, 45, and 35 min lanes from top to bottom of the time-

lapse gel (red), and the fit with the reaction-diffusion model (blue) used to extract the equilibrium or initial values of the concentrations of the monomer and dimer for (A) SAM18 and (B) SCM18.

Demonstration of the FE-MC method with hairpin structures

We also applied the FE-MC method to a hairpin structures (5'-TTGTCAT TTTTTTTTTTTTTTTTATGACTT-3') for which the folding energy has been well predicted by DNA folding algorithms^{26,27}. The hairpin had a total length of 29 nt, like HD22, and contained 5 bp in the stem and 15 nt in the loop. The folding energy predicted by the DINAMelt web server²⁶ was 2.16 kcal/mol (or 9.24 kJ/mol) under ambient conditions.

Table S1. DNA sequences used to form hairpin SAMs with N_d ranging from 18–24 bp.

	<i>Strand A (5' to 3')</i>	<i>Strand B (5' to 3')</i>
$N_d = 24$	ACG TGA GAG CAG TTGTCAT TTTTTTTTTTTTTTT ATGACTT GAC ATA CGA CGA	CTG CTC TCA CGT TCG TCG TAT GTC
$N_d = 22$	ACG TGA GAG CA TTGTCAT TTTTTTTTTTTTTTT ATGACTT AC ATA CGA CGA	TG CTC TCA CGT TCG TCG TAT GT
$N_d = 20$	ACG TGA GAG C TTGTCAT TTTTTTTTTTTTTTT ATGACTT C ATA CGA CGA	G CTC TCA CGT TCG TCG TAT G
$N_d = 18$	ACG TGA GAG TTGTCAT TTTTTTTTTTTTTTT ATGACTT ATA CGA CGA	CTC TCA CGT TCG TCG TAT

A series of SAMs containing this hairpin were prepared using the DNA sequences shown in **Table S1**, following the sample preparation procedures in Methods. Time-lapse gel images of these SAMs in each group were taken as shown in **Fig. S4A**.

Splitting of the monomer bands was observed for the SAM with $N_d = 20$, indicating that this was the critical molecular clamp length. The SAM with $N_d = 18$ (denoted as ‘SHM18’) was thus chosen for subsequent detailed analysis. Again, we obtained the initial equilibrium monomer and dimer concentrations by fitting the time-lapse gel profiles of SHM18 (as shown in **Fig. S4**) using the reaction-diffusion model. The mobilities for both monomers and dimers were extracted using a linear fitting for vertical positions of the corresponding bands as a function of time and found to be $V_m = 9.85 \times 10^{-2}$ cm/min and $V_d = 7.86 \times 10^{-2}$ cm/min. D_m and D_d were found to be 2.0×10^{-5} cm²/min and 1.9×10^{-5} cm²/min, respectively. We found that $C_{m0} = 0.53$ μ M, $C_{d0} = 0.735$ μ M, $k_1 = 0.0005$ μ M⁻¹ min⁻¹, and $k_2 = 0.0190$ min⁻¹ generated the best fitting results for the gel profiles at 55, 45, and 35 min, as shown in **Fig. S4B**. The internal energy of the SHM18 was computed using Eq. (1) to be $\Delta G_{SHM18} = 9.39 k_B T$.

Since the total length of the chosen hairpin was 29 nt, we used the same SCM18 as in the main text as a control. We measured an internal energy for SCM18 of $\Delta G_{SCM18} = 5.62 k_B T$. Therefore, the folding energy of the hairpin could be computed by $\Delta G_{fold}^{hairpin} = \Delta G_{SHM} - \Delta G_{SCM} = 3.77 k_B T$, or 9.05 kJ/mol, which was highly consistent with the value predicted by the DINAMelt web server.

Figure S4. (A) Time-lapse gel pictures for SAMs with molecular clamps of $N_d = 18$, 20, 22, and 24. (B) Intensity profiles of the 55, 45, and 35 min lanes (from left to right) of the time-lapse gel (red), and the fit with the reaction-diffusion model (blue) used to extract the equilibrium or initial values of the concentrations of the monomer and dimer for the hairpin and the molecular clamp of $N_d = 18$.

- 1 McKeague, M. & DeRosa, M. C. Challenges and Opportunities for Small Molecule Aptamer Development. *Journal of Nucleic Acids* **2012**, 20, doi:10.1155/2012/748913 (2012).
- 2 Qu, H. *et al.* Rapid and Label-Free Strategy to Isolate Aptamers for Metal Ions. *ACS Nano* **10**, 7558-7565, doi:10.1021/acsnano.6b02558 (2016).
- 3 Keefe, A. D., Pai, S. & Ellington, A. Aptamers as therapeutics. *Nature Reviews Drug Discovery* **9**, 537, doi:10.1038/nrd3141 (2010).
- 4 Fang, X. & Tan, W. Aptamers generated from cell-SELEX for molecular medicine: a chemical biology approach. *Accounts of chemical research* **43**, 48-57, doi:10.1021/ar900101s (2010).
- 5 Long, S. B., Long, M. B., White, R. R. & Sullenger, B. A. Crystal structure of an RNA aptamer bound to thrombin. *RNA*, doi:10.1261/rna.1239308 (2008).
- 6 Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. Stretching DNA with optical tweezers. *Biophysical Journal* **72**, 1335-1346, doi:10.1016/S0006-3495(97)78780-0 (1997).
- 7 Woodside, M. T., García-García, C. & Block, S. M. Folding and unfolding single RNA molecules under tension. *Current opinion in chemical biology* **12**, 640-646, doi:10.1016/j.cbpa.2008.08.011 (2008).

- 8 Bustamante, C., Bryant, Z. & Smith, S. B. Ten years of tension: single-molecule DNA mechanics. *Nature* **421**, 423, doi:10.1038/nature01405 (2003).
- 9 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, doi:10.1126/science.1151298 (2008).
- 10 Hummer, G. & Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Sciences* **98**, 3658, doi:10.1073/pnas.071034098 (2001).
- 11 Hummer, G. & Szabo, A. Free energy profiles from single-molecule pulling experiments. *Proceedings of the National Academy of Sciences* **107**, 21441, doi:10.1073/pnas.1015661107 (2010).
- 12 Ciliberto, S. Experiments in Stochastic Thermodynamics: Short History and Perspectives. *Physical Review X* **7**, 021051, doi:10.1103/PhysRevX.7.021051 (2017).
- 13 Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters* **78**, 2690-2693, doi:10.1103/PhysRevLett.78.2690 (1997).
- 14 Gore, J., Ritort, F. & Bustamante, C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences* **100**, 12564, doi:10.1073/pnas.1635159100 (2003).
- 15 Cohen, E. G. D. & Mauzerall, D. A note on the Jarzynski equality. *Journal of Statistical Mechanics: Theory and Experiment* **2004**, P07006, doi:10.1088/1742-5468/2004/07/p07006 (2004).
- 16 Vilar, J. M. G. & Rubi, J. M. Failure of the Work-Hamiltonian Connection for Free-Energy Calculations. *Physical Review Letters* **100**, 020601, doi:10.1103/PhysRevLett.100.020601 (2008).
- 17 Mergny, J.-L., Phan, A.-T. & Lacroix, L. Following G-quartet formation by UV-spectroscopy. *FEBS Letters* **435**, 74-78, doi:https://doi.org/10.1016/S0014-5793(98)01043-6 (1998).
- 18 Bončina, M., Lah, J., Prislan, I. & Vesnaver, G. Energetic Basis of Human Telomeric DNA Folding into G-Quadruplex Structures. *Journal of the American Chemical Society* **134**, 9657-9663, doi:10.1021/ja300605n (2012).
- 19 Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic acids research* **36**, 5482-5515, doi:10.1093/nar/gkn517 (2008).
- 20 Haq, I., Chowdhry, B. Z. & Jenkins, T. C. "Calorimetric techniques in the study of high-order DNA-drug interactions" in *Methods in Enzymology* Vol. 340, 109-149 (Academic Press, 2001).
- 21 Spink, C. H. "Differential scanning calorimetry" in *Methods in Cell Biology* Vol. 84, 115-141 (Academic Press, 2008).
- 22 Tasset, D. M., Kubik, M. F. & Steiner, W. Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *Journal of Molecular Biology* **272**, 688-698, doi:https://doi.org/10.1006/jmbi.1997.1275 (1997).
- 23 Qu, H. & Zocchi, G. The complete bending energy function for nicked DNA. *EPL (Europhysics Letters)* **94**, 18003 (2011).
- 24 Qu, H., Tseng, C.-Y., Wang, Y., Levine, A. J. & Zocchi, G. The elastic energy of sharply bent nicked DNA. *EPL (Europhysics Letters)* **90**, 18003 (2010).
- 25 Russo Krauss, I., Pica, A., Merlino, A., Mazzarella, L. & Sica, F. Duplex-quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA

- aptamer. *Acta Crystallographica Section D* **69**, 2403-2411, doi:10.1107/S0907444913022269 (2013).
- 26 Markham, N. R. & Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic acids research* **33**, W577-W581, doi:10.1093/nar/gki591 (2005).
- 27 Peek, A. S. *et al.* IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research* **36**, W163-W169, doi:10.1093/nar/gkn198 (2008).
- 28 Shunong, Y., Wolska-Klis, M. M. & Cann, J. R. Gel electrophoresis of reacting macromolecules. Rate-limited self-association. *Analytical Biochemistry* **196**, 192-198, doi:https://doi.org/10.1016/0003-2697(91)90138-J (1991).
- 29 Cann, J. R. Theoretical studies on the mobility-shift behavior of binary protein-DNA complexes. *Electrophoresis* **14**, 669-679, doi:10.1002/elps.11501401107 (1993).
- 30 Shubsda, M., Goodisman, J. & Dabrowiak, J. C. Characterization of hairpin-duplex interconversion of DNA using polyacrylamide gel electrophoresis. *Biophysical Chemistry* **76**, 95-115, doi:https://doi.org/10.1016/S0301-4622(98)00217-8 (1999).
- 31 Bustamante, C., Bryant, Z. & Smith, S. B. Ten years of tension: single-molecule DNA mechanics. *Nature* **421**, 423-427, doi:10.1038/nature01405 (2003).
- 32 Bustamante, C., Smith, S. B., Liphardt, J. & Smith, D. Single-molecule studies of DNA mechanics. *Current Opinion in Structural Biology* **10**, 279-285, doi:https://doi.org/10.1016/S0959-440X(00)00085-3 (2000).
- 33 Qu, H., Wang, Y., Tseng, C.-Y. & Zocchi, G. Critical Torque for Kink Formation in Double-Stranded DNA. *Physical Review X* **1**, 021008, doi:10.1103/PhysRevX.1.021008 (2011).