Kinetics of the Triplex-Duplex Transition in DNA

Il-Buem Lee,[†] Seok-Cheol Hong,[†*] Nam-Kyung Lee,[‡*] and Albert Johner[§]

[†]Department of Physics, Korea University, Seoul, South Korea; [‡]Department of Physics, Sejong University, Seoul, South Korea; and [§]Institute Charles Sadron, Centre National de la Recherche Scientifique, Universite' de Strasbourg, Strasbourg, France

ABSTRACT   The kinetics of triplex folding/unfolding is investigated by the single-molecule fluorescence resonance energy transfer (FRET) technique. In neutral pH conditions, the average dwell times in both high-FRET (folded) and low-FRET (unfolded) states are comparable, meaning that the triplex is marginally stable. The dwell-time distributions are qualitatively different: while the dwell-time distribution of the high-FRET state should be fit with at least a double-exponential function, the dwell-time distribution of the low-FRET state can be fit with a single-exponential function. We propose a model where the folding can be trapped in metastable states, which is consistent with the FRET data. Our model also accounts for the fact that the relevant timescales of triplex folding/unfolding are macroscopic.

INTRODUCTION

Since a triple helical nucleic acid was discovered (1), numerous biochemical studies have followed to investigate its structures and biological functions (2,3). The pyrimidine-motif triplex, named as such because the third strand is homo-pyrimidine, was the first identified and most widely studied of all triplexes. In the triplex, the third strand binds to the homo-purine strand of the homo-purine/homo-pyrimidine tract of a duplex via Hoogsteen basepairing (Fig. 1 A) (4). Thus, the building blocks of the triplex are CG*C$^+$ and TA*T triads where the asterisk (*) denotes Hoogsteen pairing and C$^+$ indicates the protonated cytosine. Thermodynamic properties of triplex have been studied by many authors (2,5,6). For example, the stability of the H-DNA (intramolecular triplex) structure depends on temperature and salt concentration as well as pH. The kinetics of triplex formation has also been studied (6–11).

A recently developed single-molecule fluorescence resonance energy transfer (FRET) technique allowed us to access the kinetics of triplex formation at the molecular level (12). As shown in Fig. 1 B, a triplex is thought to be obtained in vivo from a denatured duplex. In our study, we used a DNA molecule designed to form a triplex and report its formation by the change in FRET efficiency (Fig. 1 A). The molecule consists of a purine strand (AAG)$_5$ and a longer strand, including a short flexible linker flanked by two pyrimidine sequences (TTC)$_5$. The pyrimidine tract at 5′ side is complementary to the purine strand and the 5′ and 3′ sides are mirror-symmetric to each other. The long pyrimidine strand folds back like a hairpin conformation when a triplex is formed. Here we mainly focus on the kinetics of triplex formation of the molecule described above. By tracking single-molecule FRET signals from donor and acceptor dyes, the event of triplex formation can be detected (12). It has been shown that the Hoogsteen

pair in the CG*C$^+$ triad is more stable than the one in the TA*T triad at low pH. It is mainly because the protonated cytosine establishes additional electrostatic stabilization of the negatively charged backbone in the acidic condition (13).

However, in high pH conditions, cytosine is deprotonated and the CG*C$^+$ triad becomes unstable and no longer available. One of the factors determining the stability of triplex is the pK$_a$ values of cytosine. The pK$_a$ value of the cytosines in the third strand were estimated by various experimental methods (ultraviolet absorption spectroscopy, circular dichroism spectroscopy, fluorescence, and NMR) (3,14) and also calculated by several authors theoretically (15,16). These studies reported that the pK$_a$ value of a cytosine strongly depends on its location and the neighboring chemical environment. For example, the pK$_a$ value of a cytosine in a single-stranded DNA (ssDNA) lies between 5.2 and 5.5 while in a triplex it does between 8.4 and 9.5. For the cytosine with a given pK$_a$ value, the probability to be protonated is $1/(10^{pH-pK_a} + 1)$.

In neutral conditions (e.g., pH = 7.5), the cytosine in a ssDNA remains deprotonated (pK$_a$ < 6) but the cytosine in the triplex (pK$_a$ > 7.5) is likely to be protonated. When the single strand and the duplex are separate, the cytosines become unprotonated due to their small pK$_a$ value on a single strand. Although the protonation in general is a very fast process under proper pH-versus-pK$_a$ conditions, the cytosine protonation requires a longer time at moderately acidic pH. To acquire a proper pK$_a$ value for protonation, the ssDNA must first align properly in the major groove of the duplex. Thus, in the first step of triplex formation, Hoogsteen pairs are established only in TA*T triads. The triplex is later further stabilized by forming Hoogsteen pairs in CG*C$^+$ triads by protonated cytosines.

Another element affecting the folding kinetics is electrostatic interactions. A ssDNA (not protonated) typically also carries negative charge like the duplex backbone. Although electrostatic interactions are screened by salt ions in

FIGURE 1 (*A*) Triplex-forming DNA construct composed of a long mirror-repeat homo-pyrimidine sequence and a homo-purine sequence complementary to the 5′ side of the former. When the triplex folds, the energy transfer will occur between a donor (*green*, *left-hatched*) and an acceptor dye (*red*, *solid*). (*B*) Typical triplex-forming situation from a duplex DNA. Watson-Crick pairs (*dashes*). Hoogsteen pairs (*stars*). 3′ ends of the strands (*gray arrowheads*). In panels *A* and *B*, the triplex-forming (*third*) strand is shown (*red base letters* and *red line*, respectively).

a length-scale larger than the Debye screening length κ^{-1}, there is electrostatic repulsion between the strands at local scales. As we discuss later in detail, a single-stranded tail linked to the duplex forms a loop and returns to the duplex with its free end. To establish such a contact, it is expected to overcome a large energy barrier. Below, we mainly discuss the kinetics of triplex folding/unfolding in the neutral pH condition, based on the analysis of the FRET time traces. In the neutral pH condition, the long-tailed molecules switch back and forth within macroscopic times between folded and unfolded states, corresponding to the high- and low-FRET states, respectively. The dwell-time distribution of the high-FRET state is characterized by the statistics of a triplex trapped in metastable states with different macroscopic lifetimes. We conceive a model accounting for the macroscopic timescales of triplex formation and the mechanism of triplex folding. In the model, we propose that triplex folding can be established in the following three steps: (i) Approach of the free end of the single-stranded tail to the duplex across the energy barrier, (ii) Fast zipping by establishing Hoogsteen pairs, and (iii) Terminal loop reshaping/spreading of the free end into the distal part of the duplex. This picture is well supported by experimental results shown below. This observation may have interesting biological implications. As suggested previously (17–19), the triplex may interfere with RNA transcription as a roadblock to RNA polymerase. Although the triplex should be marginally stable in physiological pH, the slow kinetics still makes the stalling of RNA polymerase highly plausible because it can hamper the progression of RNA polymerase kinetically.

EXPERIMENTAL METHODS

The experimental details were already given in Lee et al. (12). Here we briefly summarize the samples and experimental scheme. We purchased all the oligonucleotides from Integrated DNA Technologies (Coralville, IA). The molecular construct for H-DNA consists of the strands 5′ AAG AAG AAG AAG AAG (Cy5) TGG CGA CGG CAG CGA (Bio) 3′ and 5′ TCG CTG CCG TCG CCA CTT CTT CTT CTT CTT TTT TCT TCT TCT TCT TCT TC (Cy3) 3′, where *Bio* stands for a biotin label for surface immobilization, and *Cy3* and *Cy5* are a donor and an acceptor dye for FRET measurements, respectively. In HDNA-6 and HDNA-9 samples, the donor dye (Cy3) is located 6 nucleotides (nts) and 9 nts away from the 3′ end of the second strand, respectively. The first and second strands were dissolved in T50 buffer (10 mM Tris-HCl, 50 mM NaCl, pH = 7.5) and incubated above the melting temperature (>90°C), and slowly cooled down so that the molecules were properly hybridized to form the final constructs as illustrated in Fig. 1.

To detect the folding and unfolding of the molecular construct in real time, we utilized the single-molecule FRET technique. The detail of the technique is given in Joo and Ha (20). To perform FRET measurements under various pH values, we used 50 mM MES (for pH 6.5), 50 mM HEPES (for pH 7.5), or 50 mM Tris-HCl (for pH 8.5). For the measurements with HEPES, we also varied Na^+ concentrations (26, 50, and 100 mM) using 26 mM NaOH and appropriate concentrations of NaCl for the rest of Na^+. These buffers also contained trolox, glucose, and gloxy as recommended for single-molecule fluorescence experiments (20).

RESULTS

Single-molecule FRET experiments on the H-DNA molecule

We have performed single-molecule FRET experiments on the DNA molecule shown in Fig. 1 *A*. From this DNA molecule, a high FRET efficiency should result when the molecule is folded into a triplex and a low FRET efficiency should result when it is unfolded (duplex + ssDNA). When the pH was low (pH < 7), only the high FRET efficiency (~0.9) was observed, indicating that a stable triplex structure was formed (Fig. 2 *A*). In the regime of high pH (pH > 8), only the low FRET efficiency (~0.1) was observed (Fig. 2 *C*). In the neutral pH condition, both low and high FRET efficiencies existed (Fig. 2 *B*). The coexistence of the two FRET states implies frequent interconversions between the two states, as described in the third subsection of this section.

Calibration of the high-FRET states

To calibrate the high-FRET states, we prepared DNA molecules with three different donor positions. For this test, we carried out experiments at a low pH value (pH = 6.5) for stable binding and the triplex was further stabilized by divalent cations ([Mg^{2+}] = 10 mM). Under these conditions, the terminal loop is known to be tight to the minimal size (5 nts) dictated by molecular-bending/torsional potential and the remaining 15 nts of the ssDNA go into the triplex. Once the triplex is folded completely, the donor and acceptor dyes would be separated by a small
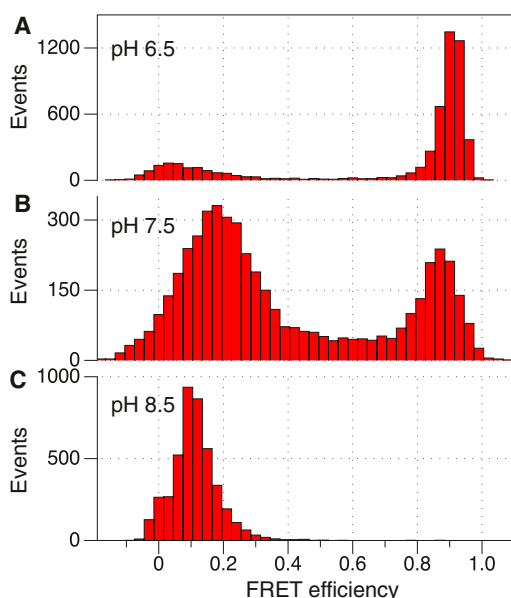
FIGURE 2   FRET efficiency histograms at various pH conditions. (*A*) 50 mM MES at pH 6.5, (*B*) 50 mM HEPES at pH 7.5, and (*C*) 50 mM Tris-HCl at pH 8.5. A triplex stabilized at low pH exhibits high FRET values.
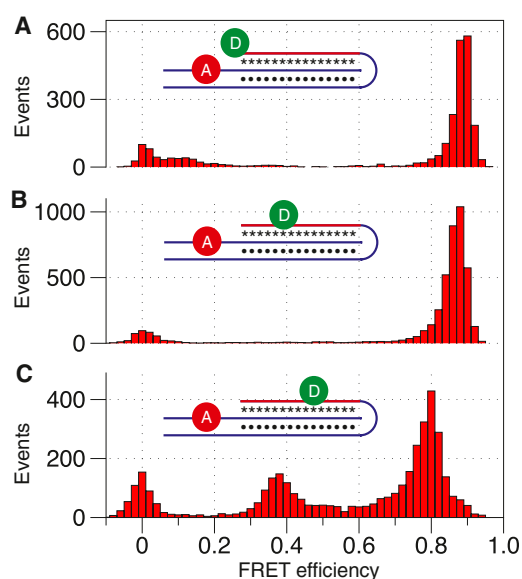

FIGURE 3   Calibration of the FRET efficiency with respect to the location of the donor dye. (*A*) The donor is next to the acceptor, (*B*) 6 nts away (HDNA-6), and (*C*) 9 nts away (HDNA-9). (*Inset*) Molecular arrangement for each measurement. The high FRET peak shifted according to the shift of donor position.

distance (1 nt or less), 6 nts, and 9 nts (see Fig. 3). As shown in Fig. 3, the FRET efficiencies for Fig. 3, *A–C*, were 0.9, 0.88, and 0.8, respectively. These values shall be considered as high FRET although the dye locations differ by as large as 9 nts. When the donor-acceptor pair was 15 nts away, the FRET efficiency peak appeared at ~0.35 (data not shown).

We note that for the molecule shown in Fig. 3 *C*, the FRET efficiency histogram showed another peak at ~0.35 and implied the donor dye was positioned approximately at the end of the duplex stem. It is presumably because triplex folding can be occasionally incomplete due to the steric hindrance of the bulky dye, which prevents further spreading of the triplex. When the dye is stuck at the entry of the duplex, the triplex stem length would be 9 nts long and the FRET efficiency of the state exhibited a mid-to-low FRET. On the contrary, for the other intermediate dye position (see Fig. 3 *B*), we did not observe any peak in the range of mid-to-low FRET values. This indicates that the triplex with 6 nts long is not stable enough. This observation is remarkably similar to the recent report regarding the minimum number (≈ 7) of basepairs required for stable duplex formation (21). In summary, these series of calibration experiments showed that the high-FRET state in the end-labeled molecule could be a set of triplex states with different numbers of Hoogsteen pairs.

Kinetics of the triplex transition

In the neutral pH condition, both low and high FRET efficiencies existed (Fig. 2 *B*). We further explored the details

of the phenomena. Fig. 4 shows the FRET efficiency histograms (*first column*), the dwell-time distributions of the low- and high-FRET states (*second* and *third columns*), and representative FRET time traces (*fourth column*) for three different Na^+ concentrations (Fig. 4 *A*, 26 mM Na^+; Fig. 4 *B*, 50 mM Na^+; Fig. 4 *C*, 100 mM Na^+) in the neutral pH condition. The FRET efficiency histograms in the first column are constructed by collecting all the time-trace data used to draw the dwell-time histograms in the second and third columns. Thus, we can avoid the counts from donor-only molecules and any contribution from noninterconverting molecules. The time-trace data indicate that the conformation switched back and forth between the folded triplex and an unfolded structure. From the salt titration experiment, it is clear that the population of the states depends on the concentration of monovalent cation, suggesting that the electrostatics has a quantitative influence on the folding/unfolding transition. The salt effect on the transition will be described in the next subsection.

In this section, the kinetics of the transition is mainly described. The typical switching time between the folded and unfolded states was on the order of seconds (see Fig. 4, *A*2, *A*3, *B*2, *B*3, *C*2, and *C*3). This is puzzling because this timescale is much longer than the typical timescales (approximately nanoseconds) known to be required for the major configurational reorganization of biological molecules (16). Our data suggest that a considerable energy barrier is associated with the triplex folding and unfolding transitions. This macroscopic timescale is also reminiscent of the timescale required for DNA hybridization reported

FIGURE 4   The conformational distribution of the triplex-forming molecule and the (un)folding kinetics under various salt concentrations. Here pH was fixed to be 7.5 by 50 mM HEPES. The data in the first row (A, red, right-hatched) were obtained with $[Na^+] = 26$ mM. The data in the second (B, blue, left-hatched) and the third (C, green, cross-hatched) rows show similar data with $[Na^+] = 50$ and 100 mM, respectively. The first, second, third, and fourth columns show the FRET efficiency histograms, the dwell-time histograms of the low- and the high-FRET states, and representative FRET time traces of each corresponding condition, respectively. The time constants obtained for this set of data and other experimental data sets are summarized in Table S1. In the dwell-time distributions, we used a bin with the same size (0.1 s).

recently (21). In the first subsection in the Discussion, we provide a physical picture of the triplex transition and an estimate for the energy barriers which are, in part, of electrostatic origin.

The dwell times of the high- and low-FRET states are analyzed. Interestingly, the dwell-time distributions for the two states are qualitatively different under low salt conditions. The gray curves in the second and third columns show fits to the dwell-time histograms. For the dwell-time of the high-FRET state (third column), we used a double-exponential fit and each curve component of the fit is shown in black dashed and dotted lines on the graph. In 26 mM Na^+, for instance, the dwell-time histogram of the low-FRET state can be described as a single-exponential decay with a characteristic time $t_{low} = 1.75$ s. On the contrary, the dwell-time distribution of the high-FRET state requires at least two timescales, $t_{h1} = 0.15$ s and $t_{h2} = 0.85$ s for 26 mM Na^+. See Table S1 in the Supporting Material for time constants for all of our dwell-time distribution data. The results shown in Fig. 4 correspond to the data labeled Exp.1 in the table.

From the fact that the dwell times for folding and unfolding transitions are comparable, the free energy difference is only of the order of a few k_BT. To be concrete, we calculated the average dwell time $\bar{t}$ for the high-FRET and low-FRET states from the data shown in Fig. 4, A2 and A3. From the average dwell times, we expect the ratio of the statistical weights to be $\bar{t}_{high}/\bar{t}_{low} \sim 0.82/2.13 \sim 0.38$. A similar value (~0.36) can be also obtained by the ratio of the areas under the high-FRET and the low-FRET peaks in the FRET efficiency histograms (Fig. 4 A1). See Table S2 for more details and other cases.

We already mentioned that it is necessary to fit the dwell-time distribution of the high-FRET state with, at least, a double-exponential function. To characterize the quality of fitting in a more systematic manner, we applied the χ^2 goodness-of-fit test. We calculated the reduced χ^2 values for single-exponential and double-exponential fits for the high-FRET and low-FRET states, respectively. The results are summarized in Table S3. From the markedly different χ^2 values, the analysis indeed proved that double-exponential, not single-exponential, fits should be used to describe the dwell-time distributions of the high-FRET states under low salt conditions. Triple-exponential fits yielded only negligible improvement in fitting (data not shown). For 100 mM Na^+ data, we could, however, get a fit of comparable quality with a single-exponential curve.

As noted here, another point of this article is to understand the origin of the multiscale statistics. We attribute the existence of multiple timescales to the presence of metastable folded states. In the Discussion, we discuss the folding/unfolding mechanism which manifests such metastability arising from slow transitions where terminal loop rearrangements are involved.

Ionic effect on the triplex folding/unfolding transition

Due to the electrostatic repulsion, the single strand would make a contact with the duplex DNA rather with its free end. The single strand also experiences electrostatic stiffening. If electrostatic interactions are strongly screened (under high salt concentration), the expected energy barrier

would be smaller and consequently, the dwell time of the low-FRET state should be reduced accordingly. Fig. 4 manifests such a mechanism of the triplex folding. As expected, with more Na^+ present, the majority of the molecules shifted from the low- to the high-FRET state. With $[Na^+] = 100$ mM, the major peak was located at the high FRET value. This indicates that the statistical weight of the folded triplex increased with increasing salt concentrations. We also analyzed the corresponding dwell-time distribution as described. As shown in Fig. 4, $A2$, $B2$, and $C2$, the characteristic time for the folding transition (the dwell time for the low-FRET state) was indeed shortened as the salt concentration increased, which agrees to the fact that the electrostatic repulsion is screened while the characteristic time for the unfolding transition (the dwell time for the high-FRET state) was nearly independent of the salt concentration.

DISCUSSION

In this section, we provide the estimate of energy barrier associated with the macroscopic timescales of triplex folding/unfolding to account for the kinetics data shown in Fig. 4. The single strand can be considered as a model polyelectrolyte (22) and behaves as a semiflexible chain with a persistence length l_p dictated by screened electrostatics. Below, the kinetics will be discussed for the data obtained in the buffer containing 50 mM HEPES and 26 mM Na^+. In this condition, there is no small anion such as Cl^- but a bulky HEPES anion, and thus Na^+ is the only ionic species to participate in efficient screening of DNA molecules. With the contribution of the cation alone, the Debye screening length κ^{-1} is ≈ 2.5 nm. If the resulting electrostatic persistence length is larger than κ^{-1}, we may apply Odijk's formula (23), $l_p = l_0 + l_B \lambda^2 \kappa^{-2}/4$ with l_B the Bjerrum length, $\lambda = 1/l_B$ the charge per unit length, and $l_0 \sim 1$ nm the bare persistence length of a ssDNA. We estimate the persistence length of the single strand to be $l_p \sim 3.3$ nm, which lies in the valid regime for the Odijk's formula. For larger salt concentrations (e.g., 100 mM) considered experimentally, the screening length and persistence length are further reduced to $\kappa^{-1} = 1.47$ nm and $l_p = 1.7$ nm and thus the Odijk's formula still applies. We thus consider the ssDNA as a semiflexible chain.

Electrostatics and terminal loop conformation

The diffusion time for a single polymer strand of length of 10 nm is of approximately nanoseconds. The limiting step for folding is the crossing of the energy barrier against penetrating into the duplex groove. We first calculate the shape of the single-stranded tail of the molecule as if it would adsorb on the central axis of the duplex stem. In this way, the problem is scale-free and easier to handle analytically. To establish chemical bonds, we demand that the single strand

linked to the duplex should align parallel with the same central axis (Y). The adopted optimal configuration has a question-mark shape (see Fig. 5 A). As we consider the lowest energy shape for the ssDNA that bends back onto the duplex, this can be obtained with the free-end of the ssDNA tail held by a constraining force μ normal (X) to the duplex. The energy to be minimized is the standard form for a wormlike chain under external force (24,25),

$$\varepsilon = \int_0^S ds \frac{1}{2} k_B T l_p \left(\frac{d\theta}{ds}\right)^2 - \mu \int_0^S ds \cos\theta(s),$$

where S is the loop length and θ the angle between the tangent to the ssDNA and the normal pointing away from the duplex (see Fig. 5). The sought shape satisfies the first integral,

$$\frac{1}{2} k_B T l_p \left(\frac{d\theta(s)}{ds}\right)^2 - f \cos\theta(s) = \text{const}, \qquad (1)$$

with $\mu = -f$. We impose that the ssDNA leaves the duplex along the Y axis ($\theta = -\pi/2$) and comes back parallel to the duplex ($\theta = \pi/2$). It is convenient to introduce the dimensionless parameter

$$\sigma = s\sqrt{\frac{f}{(k_B T l_p)}}.$$

The origin of the curvilinear abscissa σ (or s) is shown in Fig. 5 A. The corresponding curve is an Euler elastica which
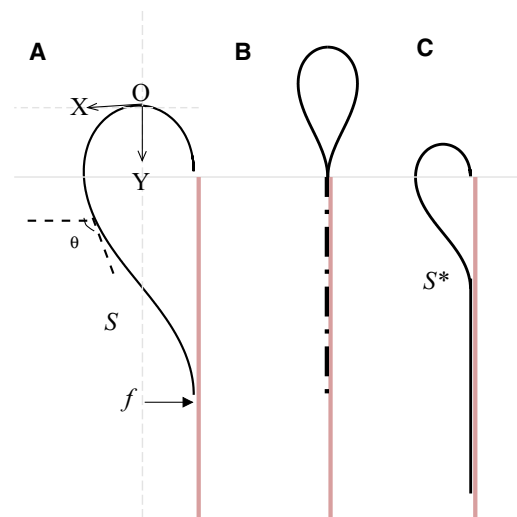

FIGURE 5 (A) Optimal shape of the single strand of length S at the moment of contact. (B) Zipped configuration where the zipped section is indicated (dot-dashed line). (C) Optimal shape of the single-stranded loop at equilibrium with loop length $S*$. The character O represents the origin of the XY coordinate system. In panels (B and C), the bound ssDNA strand is represented as if absorbed on the centerline of the duplex.

can be described in terms of the Jacobi functions SN, DN, and CN (25):

$$\sin\left(\frac{\theta}{2}\right) = \sqrt{m}\mathrm{SN}(\sigma, m),$$

$$\cos\left(\frac{\theta}{2}\right) = \mathrm{DN}(\sigma, m). \tag{2}$$

The parameter m is linked to the orientation θ_0 of the ssDNA at its inflection point through $m = \sin(\theta_0/2)^2$. The coordinates $(x(s), y(s))$ are calculated upon integration of $\cos\theta$ and $\sin\theta$ using the half angle formulas and Eq. (2).

$$X(\sigma) = -\sigma + 2E(\sigma, m),$$
$$Y(\sigma) = 2\sqrt{m}(1 - \mathrm{CN}(\sigma, m)), \tag{3}$$

where the reduced coordinates $X = x(\sigma/s)$ and $Y = y(\sigma/s)$ were introduced and E is an incomplete Elliptic integral. Geometrical constraints impose the value of the parameter $m = 0.8261$, which corresponds to $\theta_0 \approx 3\pi/4$. The shape described by Eq. 3 has a length $\sigma_S = \sigma(S) = 4.64$. This fixes the scaling factor through $\sigma(s) = s(\sigma_S/S)$ and hence the force $f = k_B T\, l_p(\sigma_S/S)^2$. The dimensionless distance between loop-ends projected on the duplex axis is ~2.2 corresponding to the actual distance of $2.2S/\sigma_S$ and the loop stores an excess length of ~$(\sigma_S - 2.2)S/\sigma_S$ as compared to its end-to-end distance. Integrating the bending energy along the loop, we find the elastic energy stored in the single strand to be $\epsilon_{loop} = 14l_p/S$ in $k_B T$ unit. The persistence length depends on the salt concentration and is typically $l_p = 2\sim5$ nm for ssDNA.

In the experiments, $l_p \approx 3.3$ nm and $14l_p/S$ could be ~4.6 $k_B T$ for the first touching loop and more for the smaller loop in equilibrium. In our experiments, the inferred electrostatic screening length (~2.5 nm) is approximately one-quarter of the contour length of the single-stranded tail (~10 nm) and the electrostatic screening is incomplete. The actual electrostatic energy of the loop is expected to be somewhat larger than the value estimated above (4.6 $k_B T$) with the assumption of strong screening (a further correction may come from the fact that the continuous wormlike chain model is not adequate at these short scales, i.e., 10–20 nts). In the experiment, the ionic strength influenced the FRET signals but not so dominantly. A large increase of the ionic strength only moderately favored the high-FRET state and reduced the dwell time of the low-FRET state by, say, a factor of 6 only. This result agrees to our estimate that the electrostatic penalty would be a few $k_B T$.

Equilibrium considerations

At equilibrium, the actual loop length S^* is smaller than S and the remaining length $S - S^*$ enters the triplex. There is a small loop at the terminus of the duplex (terminal loop) because a sharp turn of the single strand to adsorb in the groove of the duplex is not allowed due to its (electrostatic and steric) rigidity. In a simple continuous model, the total free energy reads $F = E - \epsilon(S - S^*)/l$, where ϵ designates the average free energy gain per nucleotide (of length l) in the triplex and a contribution of shape fluctuations has been ignored. The optimal shape of the ssDNA in the ground state is expected to be the same as the first contact loop given by Eq. 3, albeit with the smaller loop length S^*. The elastic bending energy is $\epsilon^*_{loop} = 14l_p/S^*$. At the equilibrium length of the loop S^* ($< S$), the derivative of the loop energy should compensate for the (average) adsorption energy ϵ of the single strand. The equilibrium loop size S^* can be obtained by minimizing the free energy $F = 14l_p/S^* - \epsilon(S - S^*)/l$. Hence, we get $\epsilon = 14l_p l/S^{*2}$. The total free energy of the adsorbed state should be negative, the open state being taken as a reference.

Interestingly, minimizing the energy shows that the critical triplex stem length is as long as the total loop length at equilibrium: $S^*_c = S/2$. In other words, the single-stranded loop can be stabilized against opening if the adsorbed length is larger than the length of the loop ($S > 2S^*$). In the experiment, the total number of nucleotides in the single-stranded tail was 20 and thus the loop at the duplex terminus can be, at most, 10 nts long and the triplex stem is at least 10 nts long. Moreover, the persistence length is related to the binding energy ϵ by $l_p = \epsilon l n^2/14$, where $n = S^*/l$ is the number of nucleotides in the loop.

As reflected from the fact that the system spends comparable times in the low- and high-FRET states, the triplex is marginally stable in our experiment. The estimate of the average binding free energy ϵ per base is obtained from the equality of terminal loop and triplex length (criticality), with $S^* \approx 5$ nm, $l = 0.5$ nm, and $l_p \approx 3.3$ nm (see the beginning of this Discussion section), and we obtain

$$\epsilon = \frac{14l_p l}{S^{\star 2}} \approx 1k_B T. \tag{4}$$

Because the bending contribution is found rather soft, we anticipate that the fluctuation δS^* around S^* in equilibrium is important. Equating the second-order derivative of the free energy F with the thermal energy, we get $(\delta S^*)^2 = (S^*)^3/(14l_p)$, which corresponds to a fluctuation $|\delta S^*|$ of ~3 nts. This happens to be just enough for the system to match the three-base periodicity of both the duplex and the ssDNA. Note that under acidic pH conditions, ϵ becomes much stronger due to cytosine-based triads and the terminal loop size shrinks down to 5 nts (16). Each TA*T corresponds to a gain in free energy of $\epsilon_H \sim 8$ $k_B T$ from two extra hydrogen bonds in Hoogsteen pairs (4). Most of the energy from H-bonds must be compensated by electrostatic repulsion and local distortions, and therefore the energy cost in total must be as high as $\Delta \sim 7$ $k_B T$ on average per base. This value may seem high, but it is not unreasonable because each base loses some

degrees of freedom (~1–2 k_BT), the strong electrostatics in the groove where the dielectric constant is four times smaller than in bulk water (16) costs $\epsilon_{el} \sim$ 1–2 k_BT per base, and H-bonds imply some distortion of the molecules, including changes in stacking (4–5 k_BT or more is reasonable).

Mechanism of triplex folding/unfolding

In the next subsections, we mainly discuss the kinetics in the strongly screened regime. To initiate the formation of the triplex, the single strand first has to bend over and make a contact with the duplex by its free end, which is the energetically most favorable configuration. The energy barrier due to the elastic/electrostatic bending energy is ~$14l_p/S \sim$ 4–5 k_BT (roughly, $S = 3l_p$). Once confined in the duplex groove, nucleotides of the ssDNA tail are strongly hindered, which gives an additional contribution to the barrier. A further contribution comes from local electrostatic interaction with the duplex. After the first triad(s) are formed, a triplex can grow through zipping. At the end of the fast zipping step, there is a remaining loop of length $(\sigma_S - 2.2)S/\sigma_S$ (see Fig. 5 B) for a complete zipping.

Our estimates show that the triplex length is about the same in the zipped state and in the ground state under neutral pH (compare Fig. 5 B and Fig. 5 C). The terminal loops also have similar sizes but quite different shapes. To reach the ground state, the loop hence has to reshape and the triplex has to shift along the DNA. The frame shift in a triplex may be quite slow. For a marginally stable triplex, the zipped state (Fig. 5 B) reaches ~10 nts from the duplex tip and exhibits high FRET as does the equilibrium state (Fig. 5 C) reaching ~15 nts from the duplex tip. We will argue that unfolding does occur from states shown in Fig. 5 B and C, which have almost the same triplex length but do not easily interconvert. Depending on the pH condition, there may be a large extra length stored in the loop at the duplex terminus, larger than the loop at equilibrium. There is a last step we call "spreading": if the loop length is larger than S^*, the terminal loop has to feed the triplex. The feeding is slower than the zipping, but typically faster than the initial binding for a short ssDNA. Again, fine tuning of the triplex length in the last stage may be slow and preempted by triplex opening. Below, we describe the processes involved in some detail.

First ssDNA/duplex binding

Starting from the most extended shape, the strand has to bend over. Because the energy barrier for looping is involved, this process should be activated. We assume that the first binding is a reaction-limiting step (26). This means that even if the end-section of the single strand is located inside the groove and adopts the orientation parallel to the duplex as required for triplex formation, there is an addi-

tional local energy barrier against binding. In practice, there is a chemical reaction rate Q or a characteristic reaction time $1/Q$ for a base on ssDNA already assumed in a proper contact with the groove of the duplex. The reaction rate is then equal to the Boltzmann weight of the activated configuration times the reaction rate Q dictated by local barriers. The hydrogen bond(s) will only be established with a certain orientation (deformation) of the bases and their environment.

Considering the first contact shape shown in Fig. 5 A, we recall that the duplex spans with the radius of $R = 1$ nm from the central axis so that some of ssDNA bases are hence located in the groove of the duplex or its very vicinity. Thereby n_t bases are confined in the groove. The trapped nucleotides give up their lateral and part of their radial fluctuations. From a geometrical estimate based on the model presented in Fig. 5 A, we conclude that there are at least five nucleotides caught in the groove. In the following discussion, we assume that the number of trapped nucleotides is $n_t = 5$ (note that, to diminish its global hindrance, some slight deformation of ssDNA is likely) and the associated penalty for trapping is ~1 k_BT per base. The free energy cost associated with the hindrance in the groove is thus estimated to be $E_g \sim 5\ k_BT$.

Writing $Z_{g/f}$ as the partition function of the groove/free state, the ratio $Z_g/Z_f \sim e^{-E_g/k_BT}$ reflects the hindrance of the n_t trapped nucleotides in the groove. The equilibrium contact probability obeys Boltzmann law and is related to its free energy penalty dF:

$$e^{-dF/k_BT} = e^{-14l_p/S}\ \frac{Z_g}{Z_f}.$$

Collecting relevant factors, the closing rate k_{cl} is found to be

$$k_{cl} = Qe^{-14l_p/S-\log\left(Z_f/Z_g\right)}.$$

The reaction rate Q is linked to the local barrier Δ by $Q = \tau^{-1}\exp(-\Delta/k_BT)$, where τ is an inverse of microscopic attempt frequency which is on the order of nanoseconds. The characteristic time

$$t_{cl} = k_{cl}^{-1} \sim \tau e^{\left(\Delta+E_g\right)/k_BT+14l_p/S}.$$

turns out to be larger than the microscopic time τ by ~7–8 orders of magnitude and is hence macroscopic. The measured closing times are slightly smaller than a second. The experimental closing kinetics was described by a single-exponential function at large (experimental) time-scales, which is compatible with the mechanism considered here. Many unsuccessful attempts of a first contact, which do not lead to folding, give rise to a single averaged folding time by sampling the distribution of the first contact time.

Zipping and spreading

After the first binding by the free end of the single strand, the zipping of the strand occurs at the expense of the single stranded loop. During the zipping, the ssDNA closes onto the duplex reducing the size of the terminal loop. For zipping, we further assume a barrier of Δ (approximately a few thermal units) against entering the binding reaction. Each closing step is almost independent of others and the total zipping time is proportional to the number of closings,

$$t_z = \frac{\tau e^{\Delta/k_B T}(2.2S/l)}{\sigma_S},$$

which is ~1 ms, below the time resolution of our FRET experiment (it is further unlikely that the FRET efficiency changes during zipping because the label attached to the single-stranded tail does not move). At the end of the zipping stage, the triplex is only as long as allowed by the original contact point and hence leads to a stem length of $2.2S/\sigma_S$, where complete zipping is assumed. From the analysis based on Eqs. 2 and 3, we estimate that the stem length is ~9–10 nts (see Fig. 5 B). According to the results in Fig. 3, the zipped configuration should exhibit high FRET. While the first contact loop (Fig. 5 A) and equilibrium loop (Fig. 5 C) have the same shape (albeit different sizes), the zipping imposes a very specific geometric constraint and hence a different shape.

Assuming the complete zipping as depicted in Fig. 5 B, the loop shape after zipping obeys Eqs. 2 and 3 with the parameter $m = 0.731$ and the bending energy is $18l_p/S_z$ for a loop of size S_z. In the subsequent process, the terminal loop can almost preserve its size but has to reshape to match the equilibrium loop. Thus, approximately five nucleotides have to desorb at the duplex tip and five nucleotides must adsorb at the distal triplex. (Here we treated the ssDNA as if absorbed on the centerline of the duplex system. In reality, the ssDNA winds along the major groove, and thus more ssDNA can be accommodated in a given triplex length. After free energy minimization, under the frozen-in first contact constraint, the resulting loop shape becomes closer to the equilibrium state, rather than to the completely zipped shape shown in (Fig. 5 B), which does not alter our conclusions).This is not an easy task: An adsorbed strand, N bases long, cannot break all hydrogen bonds with the duplex at the same time and move because of the high activation energy ($\propto N$) involved. Several weakly activated processes are at work to let a length of strand flow from one side to the other through the current triplex. The mechanism of traveling loop, described in Fig. 6, is likely to be the relevant one. A small nonadsorbed loop, a bulge of single strand, has to be injected on the terminus side and diffuse through the triplex to its end. The time needed to transfer a length of strand in this process is only polynomial in the current triplex length but both the injection of the

traveling loop and its diffusion involve some activation (also a strong $CG*C^+$ triad can hamper the diffusion of traveling loops). Finally, to approach the equilibrium configuration, traveling loops have to be injected at both ends of the triplex. The kinetics is hence dominated by small (minimal) traveling loops and the activation of bulge loops implies extra activation energy of $4\epsilon \approx 4\ k_B T$. Processes involved here are variants of the exclusion process (27). More details will be presented elsewhere. The folding process is likely to be trapped in the zipped state or in a nearby state. Unfolding proceeds from this state. Because the number of triads involved in the zipped and equilibrium state is almost identical, the lifetimes of both states are close. This matches experimental results. We hence identify the zipped state as the metastable state observed at neutral pH.

Opening of triplex

Opening of a triplex is an activated process. Many attempts are needed for a successful opening, which is then further expedited by electrostatics. When all but the last Hoogsteen pair are open (one may alternatively consider activated states with only one remaining H-bond), there remains a substantial barrier roughly amounting to the energy of the two hydrogen bonds $\epsilon_H \sim 8\ k_B T$. The state with $N - 1$ pairs open except the last hydrogen bonding is realized with a Boltzmann weight $e^{-E_b/k_B T}$, where E_b takes into account $N - 1$ binding energies $(N - 1)\epsilon$, the difference in elastic energy of the loops of this state and the ground state

$$\delta\epsilon_{loop} \sim 14l_p(1/S - 1/S^*) \sim -5k_B T,$$

the free energy cost for keeping some nucleotides in the groove E_g, and some local electrostatic cost in the groove ϵ_{el}. With $N = 10$ nts in the triplex stem, the opening time is

$$t_{op} \sim \tau e^{\left((N-1)\epsilon+\delta\epsilon_{loop}+E_g+\epsilon_{el}+\epsilon_H\right)/k_B T}$$

and

$$(N-1)\epsilon + \delta\epsilon_{loop} + E_g + \epsilon_{el} + \epsilon_H \approx (9-5+5+2+8)k_B T.$$
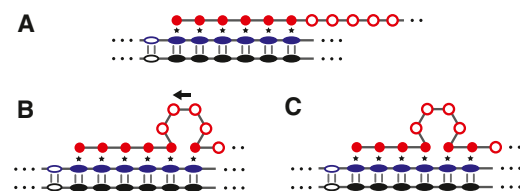

FIGURE 6 A possible mechanism for spreading. (A) Before spreading. A loop inserted is located at the first (B) and second (C) base of the triplex. (Ladder) Duplex DNA. (Red circles) Nucleotides of the third strand.

Given that $e^{19} \sim 10^8$, the opening time becomes macroscopic (remember the microscopic time τ is approximately nanoseconds). Furthermore, the opening time is also found on the same order as the closing time. These two features are consistent with experimental findings.

CONCLUSIONS

We studied the mechanism of triplex folding/unfolding using a single-molecule FRET technique. In a neutral pH, the triplex is found to be marginally stable and the molecules are switching between folded and unfolded states as revealed by alternating high and low FRET efficiencies. Our study is devoted to this regime of the neutral pH. The FRET signal is swinging between high and low values with the dwell times in the range of seconds. The average dwell time of the low-FRET state is comparable to that of the high-FRET state. In the data, the former was roughly four times larger than the latter, which indicates a small free energy difference between the two states as low as $\log 4 \approx 1.4$ in k_BT unit. This is quite remarkable, given the number of extra hydrogen bonds involved in the triplex, each pair corresponding to ~8 k_BT. There is, hence, a near-cancellation of several large contributions to the free energy. We checked experimentally that the occupancy of the two states is rather weakly affected by the salt concentration, which is consistent with the moderate contribution of large-scale electrostatics calculated above. Most of the penalty in the folded state comes from hindrance of the nucleotides of the single strand in the duplex groove and local bond distortions, including changes in stacking. The folding kinetics also reveal two intriguing features: (i) The switching time is on the order of seconds, which is macroscopic. (ii) The dwell-time distribution of the high FRET state must be fit with a multiple exponential function. We have shown here that barriers with high activation energy control the kinetics of the transitions in both directions. The estimates based on equilibrium considerations and calculations of bending energy give activation energies in the range of 17–20 k_BT associated with both folding and unfolding. Both high-FRET dwell times obtained by fitting are close, suggesting that the triplex length is almost the same in the metastable and the ground states. Under neutral pH conditions (marginal triplex stability), the zipped state typically corresponds to the same triplex length as in the ground state. The metastable state is suggested to be the zipped state, merely differing from the ground state by the terminal loop shape. The interconversion between metastable and ground states is scarce (during the lifetime of the states) and both contribute to the unfolding statistics independently, leading to a more complex dwell-time distribution of the high-FRET signal. It is also possible that the dye itself gets trapped as the single strand enters the duplex groove. Unless the dye has some specific affinity for the duplex, the molecule unfolds from this loosely bound state and consequently, this state should be (much) shorter lived.

The free energy gain per triad is on the order of the thermal energy (k_BT), similar to that of a Watson-Crick pair, and the measured lifetimes of open and closed states are reminiscent of the characteristic timescales in DNA hybridization reported by a recent work where the authors used a time resolution similar to ours (21). Much of the complexity of the folded molecule has been deliberately ignored, and some may be linked to the detailed nucleotide sequence itself or to the interplay between spatially close nucleotides (16). This could be taken into account by a numerical approach addressing specific subprocesses such as spreading or a first contact. Such a simulation approach has been applied by Pack et al. (16). Given that the activation energy is very large, any explicit simulation of the folding/unfolding process seems difficult. In that sense, a continuous coarse-grained approach as presented here provides some insights. Therefore, it seems appropriate to give timescales only in orders of magnitude (powers of 10). This is enough to answer the main puzzles raised by the experimental results and listed above.

In this study, we mainly focused on physical aspects of triplex folding/unfolding. In a neutral pH condition, the short triplex we considered is only marginally stable. With one mismatch in such a critical situation, the high-FRET dwell time is expected to be greatly reduced or the triplex might not be stable over measurable times at all. The kinetics is also expected to depend on the number of TTC triplet repeats. The unfolding is slower for longer strands because a large number of hydrogen bonds should be broken. We expect the strand length to have less influence on the closing time.

The fact that the opening time is macroscopic may have an interesting biological implication. It has been known that RNA polymerases can transcribe ~40 nucleotides of RNA per second (in single-molecule assays, the rate is ~10–20 nt/s (28,29) and stalling of RNA polymerases is more likely to happen when they are paused (29). In a more recent work, RNA polymerases proceed in a ratchet mechanism (30) and thus the DNA template needs to be opened presumably by thermal excitation. As shown in this article, the part forming a triple helical structure revealed slow kinetics with its folding and unfolding times in the range of seconds although the structure under a physiological pH condition is thermodynamically marginally stable. Consequently, RNA polymerases may proceed much more slowly through the triplex-forming sequence and have a higher chance to be arrested. This is consistent with the proposed mechanism of how the long TTC repeats can induce Friedreich ataxia in which the expression of frataxin gene is suppressed by triplex formation of the repeat sequence (18). Thus, the dynamic nature of a triplex revealed here may substantiate the proposed biological role of the triplex.

SUPPORTING MATERIAL

Three tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)01182-4.

REFERENCES

1. Felsenfeld, G., D. R. Davies, and A. Rich. 1957. Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* 79:2023–2024.

2. Frank-Kamenetskii, M. D., and S. M. Mirkin. 1995. Triplex DNA structures. *Annu. Rev. Biochem.* 64:65–95.

3. Soyfer, V., and V. N. Potaman. 1996. Triple-Helical Nucleic Acids. Springer, New York.

4. Hoogsteen, K. 1963. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* 16:907–916.

5. Plum, G. E., Y.-W. Park, …, K. J. Breslauer. 1990. Thermodynamic characterization of the stability and the melting behavior of a DNA triplex: a spectroscopic and calorimetric study. *Proc. Natl. Acad. Sci. USA.* 87:9436–9440.

6. Shindo, H., H. Torigoe, and A. Sarai. 1993. Thermodynamic and kinetic studies of DNA triplex formation of an oligohomopyrimidine and a matched duplex by filter binding assay. *Biochemistry.* 32:8963–8969.

7. Maher, 3rd, L. J., P. B. Dervan, and B. J. Wold. 1990. Kinetic analysis of oligodeoxyribonucleotide-directed triple-helix formation on DNA. *Biochemistry.* 29:8820–8826.

8. Rougée, M., B. Faucon, …, C. Hélène. 1992. Kinetics and thermodynamics of triple-helix formation: effects of ionic strength and mismatches. *Biochemistry.* 31:9269–9278.

9. Xodo, L. E. 1995. Kinetic analysis of triple-helix formation by pyrimidine oligodeoxynucleotides and duplex DNA. *Eur. J. Biochem.* 228:918–926.

10. Alberti, P., P. B. Arimondo, …, J. S. Sun. 2002. A directional nucleation-zipping mechanism for triple helix formation. *Nucleic Acids Res.* 30:5407–5415.

11. James, P. L., T. Brown, and K. R. Fox. 2003. Thermodynamic and kinetic stability of intermolecular triple helices containing different proportions of C$^+$·GC and T·AT triplets. *Nucleic Acids Res.* 31:5598–5606.

12. Lee, I. B., J. Y. Lee, …, S.-C. Hong. 2012. Direct observation of the formation of DNA triplexes by single-molecule FRET measurements. *Curr. Appl. Phys.* 12:1027–1032.

13. Sun, J. S., J. L. Mergny, …, C. Hélène. 1991. Triple helix structures: sequence dependence, flexibility and mismatch effects. *J. Biomol. Struct. Dyn.* 9:411–424.

14. Leitner, D., W. Schröder, and K. Weisz. 2000. Influence of sequence-dependent cytosine protonation and methylation on DNA triplex stability. *Biochemistry.* 39:5886–5892.

15. Petrov, A. S., G. Lamm, and G. R. Pack. 2004. Direct observation of the formation of DNA triplexes by single-molecule FRET measurements. *Biophys. J.* 87:3954–3973.

16. Pack, G. R., L. Wong, and G. Lamm. 1998. pK$_a$ of cytosine on the third strand of triplex DNA: preliminary Poisson-Boltzmann calculations. *Int. J. Quantum Chem.* 70:1177–1184.

17. Kovacs, A., J. C. Kandala, …, R. V. Guntaka. 1996. Triple helix-forming oligonucleotide corresponding to the polypyrimidine sequence in the rat α 1(I) collagen promoter specifically inhibits factor binding and transcription. *J. Biol. Chem.* 271:1805–1812.

18. Grabczyk, E., and K. Usdin. 2000. The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res.* 28:2815–2822.

19. Belotserkovskii, B. P., E. De Silva, …, P. C. Hanawalt. 2007. A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J. Biol. Chem.* 282:32433–32441.

20. Joo, C., and T. Ha. 2008. Single-molecule FRET with total internal reflection microscopy. *In* Single-Molecule Techniques. P. Selvin and T. Ha, editors. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

21. Cisse, I. I., H. Kim, and T. Ha. 2012. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.* 19:623–627.

22. Dessinges, M.-N., B. Maier, …, V. Croquette. 2002. Stretching single stranded DNA, a model polyelectrolyte. *Phys. Rev. Lett.* 89:248102.

23. Odijk, T. 1997. Polyelectrolytes near the rod limit. *J. Polymer Sci.* 15:477–483.

24. Odijk, T. 1995. Stiff chains and filaments under tension. *Macromolecules.* 28:7016–7018.

25. Lee, N.-K., A. Johner, and S. C. Hong. 2007. Compressing a rigid filament: buckling and cyclization. *Eur Phys. J. E Soft Matter.* 24:229–241.

26. O'Shaughnessy, B., and D. Vavylonis. 2003. Irreversible adsorption from dilute polymer solutions. *Eur Phys J E Soft Matter.* 11:213–230.

27. Evans, M., and T. Hanney. 2005. Nonequilibrium statistical mechanics of the zero-range process and related models. *J. Phys. Math. Gen.* 38:R195–R240.

28. Adelman, K., A. La Porta, …, M. D. Wang. 2002. Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. USA.* 99:13538–13543.

29. Forde, N. R., D. Izhaky, …, C. Bustamante. 2002. Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. USA.* 99:11682–11687.

30. Bar-Nahum, G., V. Epshtein, …, E. Nudler. 2005. A ratchet mechanism of transcription elongation and its control. *Cell.* 120:183–193.