

# Evidence for Non-Two-State Kinetics in the Nucleocapsid Protein Chaperoned Opening of DNA Hairpins

Gonzalo Cosa,<sup>§,†,‡</sup> Yining Zeng,<sup>§,†</sup> Hsiao-Wei Liu,<sup>§</sup> Christy F. Landes,<sup>§</sup> Dmitrii E. Makarov,<sup>§</sup> Karin Musier-Forsyth,<sup>#</sup> and Paul F. Barbara<sup>\*,§</sup>

Department of Chemistry and Biochemistry, Center for Nano and Molecular Science and Technology, University of Texas, Austin, Texas 78712, and Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455

Received: July 28, 2005; In Final Form: November 1, 2005

In HIV-1 reverse transcription, the nucleocapsid protein, NC, induces secondary structure fluctuations in specific DNA and RNA hairpins. Time-resolved single-molecule fluorescence resonance energy transfer was used to study NC chaperoned opening of DNA hairpins over a broader range of conditions and in more depth than in previous studies. The experiments reveal a complex mechanism for secondary structure fluctuations with dynamic processes occurring over a wide time range, i.e., ~5 to >250 ms and with the involvement of long-lived intermediates. The dynamic role of DNA loop regions and NC binding/dissociation events are discussed.

## Introduction

Nucleic acid binding proteins play a central role in many critical processes in genetics.<sup>1–3</sup> Specific examples include replication of DNA,<sup>4</sup> transcription of DNA to RNA,<sup>5,6</sup> and reverse transcription of RNA to DNA.<sup>7</sup> A well-known example is the *E. coli* single-stranded DNA binding protein, SSB, which has been observed to support replication by forming a complex with DNA in which double-stranded DNA is unwound.<sup>4</sup> The present paper is focused on another important nucleic acid binding protein, namely the nucleocapsid protein, NC, of the retrovirus HIV-1.<sup>8</sup> In addition to stabilizing the enveloped virion,<sup>9,10</sup> NC also chaperones the formation of stable nucleic acid duplexes (both RNA and DNA) in several distinct steps in the reverse transcription mechanism of the retrovirus.<sup>11–17</sup> NC destabilizes the secondary structure of the bound nucleic acid to promote subsequent annealing steps.<sup>18–22</sup> Investigating the structure and dynamics of NC/nucleic acid complexes is a critical first step for building a molecular level understanding of HIV-1 reverse transcription.

Time-resolved single molecule spectroscopy, SMS, has been applied to study nucleic acid conformational dynamics.<sup>23,24</sup> SMS can be used to unravel the dynamics of bimolecular processes that involve the interconversion of multiple structures. These studies have shown that the secondary and tertiary structures of protein/nucleic acid complexes can be highly dynamic, involving a complex distribution of nucleic acid structures that interconvert over a broad distribution of time scales. For example, the complex of reverse transcriptase with DNA was observed to adopt three distinct structures.<sup>25</sup> Single-molecule fluorescence resonance energy transfer, SMFRET, revealed a distribution of rates for the unwinding of DNA by *E. coli*

helicase.<sup>26,27</sup> Such complex processes would be difficult to characterize accurately by traditional ensemble methods alone.

RNA folding/unfolding represents a further example of dynamic complexity that can be revealed by SMS. SMFRET was used to examine heterogeneous structural dynamics in *B. subtilis* RNase P RNA.<sup>28</sup> Other single molecule RNA folding studies demonstrated that conformational transformations exhibit dynamic heterogeneity,<sup>29,30</sup> and can occur on time scales that vary by several orders of magnitude.<sup>24,31</sup> Single molecule force measurements have been used to quantify the free energy landscape for nucleic acid folding, and similar complexity has been observed.<sup>32,33</sup> It has been postulated that a strong correlation between structural dynamics and heterogeneous reaction pathways is ubiquitous.<sup>31</sup> Thus, it remains important to characterize the nucleic acid structural dynamics induced by NC.

Recently, we used SMFRET to study biotin-immobilized NC/TAR DNA complexes.<sup>20</sup> TAR DNA is a hairpin structure with complementarity to the transactivation response element, TAR RNA, and is also a component of minus-strand strong-stop DNA.<sup>20</sup> In the minus-strand transfer step of HIV-1 reverse transcription, TAR DNA must anneal to the complementary TAR RNA hairpin located at the 3' end of the viral RNA in order for minus-strand synthesis to continue.<sup>15</sup> SMFRET measurements of the 3'/5' end-to-end distance fluctuations of TAR DNA in the absence of NC revealed that TAR-DNA is in its closed form with a maximum number of Watson–Crick base pairs formed.<sup>20</sup> In contrast, SMFRET studies in the presence of NC demonstrated that NC destabilizes the secondary structure of the terminal two loop regions of the initially closed TAR DNA.

Cosa et al. proposed<sup>20</sup> that the NC/TAR DNA complex is an equilibrium mixture of closed and partially open TAR DNA/NC forms, denoted in Scheme 1 as the C and Y forms, respectively, where the loop regions are labeled by L1–L4. It is important to note that the Y terminology implies only that the Watson–Crick base pairs in the L1–L2 region have been broken due to bound NC, and does not indicate a specific secondary or tertiary structure. Indeed, considering the flexibility of the

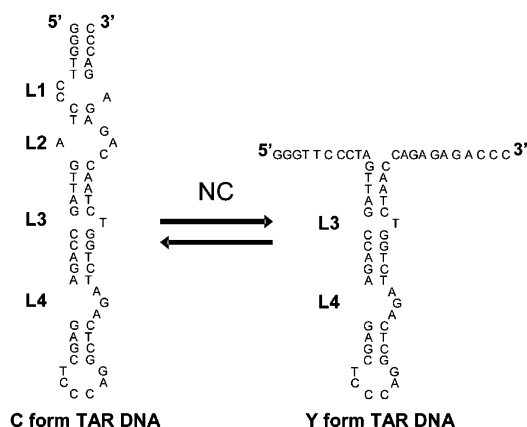
\* To whom correspondence should be addressed. E-mail: p.barbara@mail.utexas.edu.

§ University of Texas

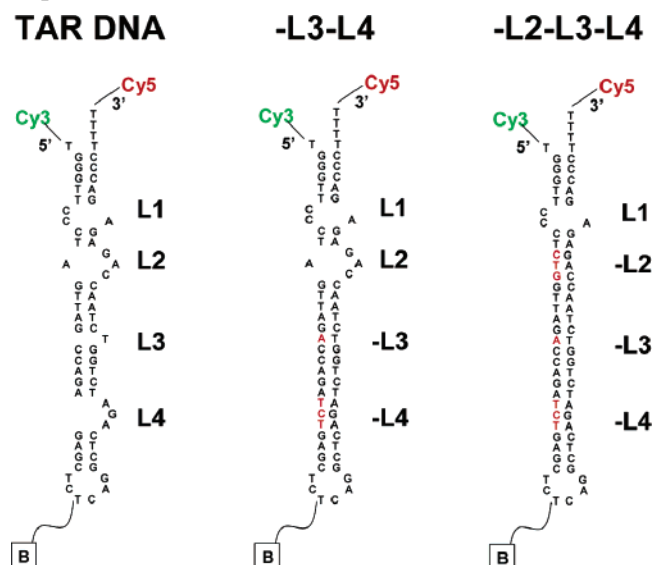
# University of Minnesota

† These authors contributed equally to this work.

‡ Present address: Department of Chemistry, McGill University, 801 Sherbrooke St. West Montreal, Quebec H3A 2K6, Canada.

**SCHEME 1: Primary and Secondary Structures of TAR DNA<sup>a</sup>**

<sup>a</sup> Left: The primary and secondary structure of TAR DNA is shown in its closed, C, form with the four loop regions labeled. Right: In the presence of NC, the secondary structure in the L1–L2 regions is disrupted, termed the Y form.

**SCHEME 2: The Primary and Secondary Structures of TAR DNA and TAR DNA Mutants Used in the Current Experiments<sup>a</sup>**

<sup>a</sup> The DNAs were immobilized on the surface via a biotin linker (noted above as “B”) attached to a dT in the hairpin loop region.

single-stranded regions and the likelihood of a broad distribution of NC binding sites, it seems likely that the Y form should in fact be structurally and perhaps dynamically heterogeneous.

Previous SMFRET experiments on TAR DNA mutants (see Scheme 2) helped assign the observed NC-induced secondary structure fluctuations to the L1–L2 region of TAR. The experiments were made at sufficiently high NC concentration to ensure saturated NC binding, i.e., 1 NC molecule/7 nucleotides (see ref 8 for review). Under these conditions the observed SMFRET dynamics were qualitatively consistent with a simple kinetically homogeneous C/Y interconversion on the 3–10 ms time scale depending on  $\text{Mg}^{2+}$  concentration.<sup>34</sup> However, Cosa et al. also observed that a small fraction of the NC/DNA complexes exhibited much slower dynamics, which were ascribed tentatively to NC/DNA complexes that were perturbed by the immobilization process.<sup>20</sup>

The current report is a more extensive and in-depth characterization of the secondary structure fluctuations of the L1–L2

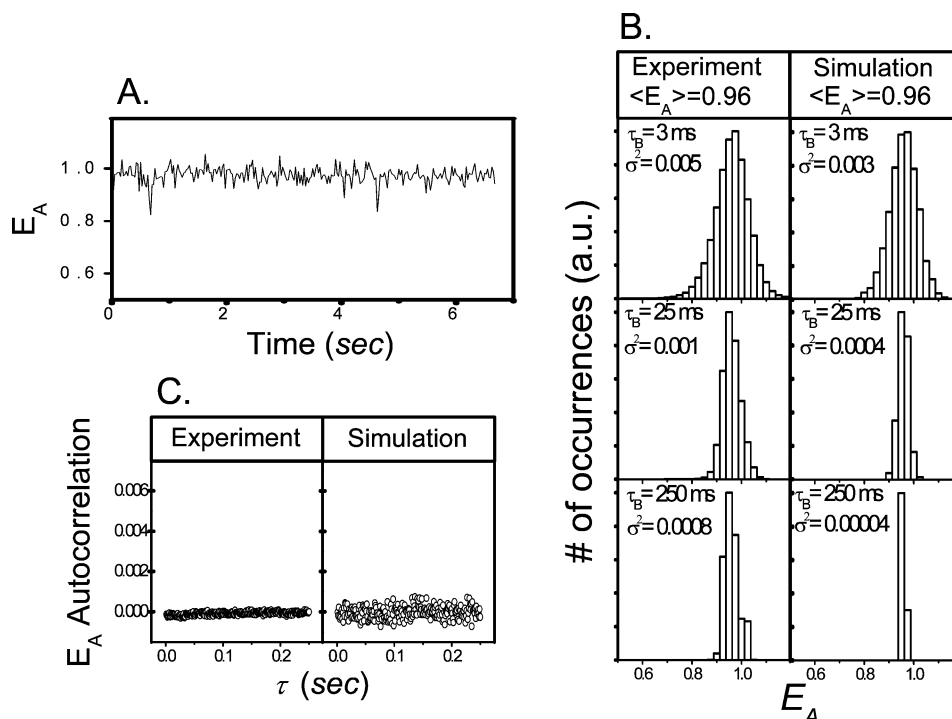
region of TAR DNA in the presence of NC. NC-induced dynamics in the L1–L2 region were measured over a larger range of NC,  $\text{Mg}^{2+}$ , and  $\text{Na}^+$  concentrations than in previous studies. These studies reveal multiple time scales for 3'/5' end-to-end distance fluctuations, ranging from single milliseconds to >250 ms. The current SMFRET results suggest that the C/Y interconversion is more dynamically complex than previously thought, and in fact involves multiple intermediate states and potentially multiple pathways. This is the first report of dynamic complexity for the NC/TAR DNA system, and may have implications for the multiple annealing pathways that have been observed in separate TAR DNA/oligonucleotide binding studies.<sup>35</sup>

**Experimental Section**

The HIV-1 NC protein used for these experiments was prepared as previously described.<sup>20,36</sup> Purified Cy3 (donor) and Cy5 (acceptor) labeled biotinylated DNA sequences were acquired from TriLink Biotechnologies (San Diego, CA). The MFOLD<sup>37,38</sup> predicted secondary structures are illustrated in Scheme 2. Dye-labeled DNA was immobilized by using either biotinylated bovine serum albumen (BSA) or biotinylated polyethylene glycol (PEG).<sup>20,35</sup> The BSA method involves assembling predrilled polycarbonate films with an adhesive gasket (Grace Bio-Labs, Bend, OR) on top of clean coverslips. Inlet and outlet ports (Nanoport, Upchurch Scientific, Oak Harbor, WA) were glued on top of the chambers. The chambers were incubated sequentially for 10 min with solutions of biotinylated BSA (Pierce Biotechnology, Inc., Rockport, IL; 2 mg/mL in distilled deionized water) and streptavidin (Molecular Probes, Eugene, OR), 0.2 mg/mL in HEPES buffer (25 mM HEPES, pH 7.3). The chambers were rinsed with distilled deionized water (50  $\mu\text{L}$ ) following each incubation step.

For the PEG immobilization, clean coverslips were treated with Vectabond/acetone 1% w/v solutions (Vector Laboratories, Burlingame, CA) for 5 min. Coverslips were then rinsed with  $\text{H}_2\text{O}$  and dried under a  $\text{N}_2$  stream. The Vectabond coated coverslips were covered with templated silicone gaskets. The exposed areas were incubated with 25% m-PEG-SPA-2000 and 0.25% biotin-PEG-NHS-5000 (Nektar Therapeutics, San Carlos, CA) in 0.1 M sodium bicarbonate (Hyclone) for 3 h. The silicone gaskets were removed, the excess PEG rinsed away, and the cover slips dried with  $\text{N}_2$ . The chamber was then treated with streptavidin and doubly labeled oligomer solution in the same manner as in the BSA method described above. Reactant solutions, which were delivered by syringe pumps,<sup>20,35</sup> contained HEPES buffer (25 mM HEPES, pH 7.3), various concentrations of  $\text{MgCl}_2$  and NaCl, and an oxygen scavenger system (2-mercaptoethanol 1% v/v (Sigma-Aldrich, St. Louis, MO),  $\beta$ -D-(+)-glucose 3% w/v (Sigma-Aldrich, St. Louis, MO), glucose oxidase 0.1 mg/mL (Roche Applied Science, Hague Road, IN), and catalase 0.02 mg/mL (Roche Applied Science)).<sup>39,40</sup>

The scanning confocal optical/data collection system used the 514 nm line from an argon ion laser (model Reliant 150m, Laser Physics, Inc., West Jordan, UT).<sup>20,35</sup> The high numerical aperture oil immersion microscope objective (Zeiss Fluor, 100X, NA 1.3) was used to focus both the excitation and emission. Donor and acceptor fluorescence was collected with two avalanche photodiodes (APD) (Perkin-Elmer Optoelectronics SPCM-AQR-15, Vaudreuil, Quebec, Canada). The intensity time trajectories were acquired with 1 ms time resolution. The collected donor and acceptor signals were corrected for background emission/noise and donor/acceptor crosstalk due to overlapping emission as previously described.<sup>20</sup> Also,  $\sim 1$  ms



**Figure 1.** SMFRET of -L3-L4 TAR DNA. (A) A representative single molecule  $E_A$  trajectory, with a bin time,  $\tau_B$ , of 25 ms. (B) Experimental  $E_A$  values presented in the form of ensemble  $E_A$  histograms with the data binned into different  $\tau_B$  (left panels). Monte Carlo simulated ensemble  $E_A$  histograms, based on a single-state system, are included for comparison (right panels). (C) The 1-ms experimental ensemble  $E_A$  trajectory is presented in the form of an autocorrelation (left panel). The simulated curve is shown for comparison (right panel).

blinking events (acceptor reversible photobleaching) were filtered from the data. Analysis of unfiltered data verified that the filtering process did not significantly distort the FRET dynamics on the time scales discussed in this paper. The corrected donor and acceptor intensities,  $I_D(t)$  and  $I_A(t)$ , respectively, can be used to calculate directly the time trajectory of the apparent FRET efficiency,  $E_A(t)$  as in:

$$E_A(t) = \frac{I_A(t)}{I_A(t) + I_D(t)} \quad (1)$$

$E_A(t)$  is related to the actual FRET efficiency,  $E_{\text{FRET}}(t)$ , by the inclusion of the dye quantum efficiencies,  $\phi_i$ , and detector quantum efficiencies,  $\eta_i$ , according to:

$$E_{\text{FRET}}(t) = \frac{I_A(t)}{I_A(t) + I_D(t) \frac{\phi_A \eta_A}{\phi_D \eta_D}} \quad (2)$$

In the case of the current experimental setup, it was determined that  $E_A(t) \sim E_{\text{FRET}}(t)$ .  $E_A(t)$  autocorrelation curves,  $E_A$ , Autocorrelation, were obtained by using the single molecule cross correlation curves, as detailed in a previous study,<sup>20</sup>

$$E_A \text{ Autocorrelation}(\tau) = C(\tau) \left( -\frac{\langle I_D \rangle}{\langle I_A \rangle} \right) \quad (3)$$

$$C(\tau) = \frac{\langle \delta I_D(t) \delta I_A(t + \tau) \rangle}{\langle I_D \rangle \langle I_A \rangle} = \frac{\langle I_D(t) I_A(t + \tau) \rangle}{\langle I_D \rangle \langle I_A \rangle} - 1 \quad (4)$$

The data were fit in most cases to a single-exponential decay, with autocorrelation lag time, amplitude, relaxation rate, and relaxation lifetime represented by  $\tau$ ,  $A_{AC}$ ,  $\lambda$ , and  $\tau'$ , respectively.

For the simulations reported here, the donor and acceptor signals,  $I_D(t)$  and  $I_A(t)$ , were simulated by using the kinetic Monte Carlo (KMC) method (for applications of KMC to FRET systems see, e.g., ref 41). If, for example, the system switches between states 1 and 2 with the FRET efficiencies  $E_1$  and  $E_2$ , respectively, then the simulation proceeds as follows. First KMC is used to generate a stochastic “trajectory” switching between 1 and 2 according to the prescribed rate constants for the  $1 \rightarrow 2$  and  $2 \rightarrow 1$  transitions. While the molecule remains in one of the states, say state 1, the donor and acceptor signals are Poisson processes, with the probability of photon arrival per unit time equal to  $I_0(1 - E_1)$  and  $I_0 E_1$ , respectively. Here  $I_0$  is the observed intensity (the number of photons per unit time) from the donor in the absence of FRET. The above formula assumes the same photodetection efficiency for both channels, which is the case for the current system. Again, a KMC procedure was used to generate random lists of photon arrival times obeying the above Poisson processes. To account for the crosstalk between the donor and the acceptor channels, whenever the algorithm generated a donor photon arrival event, it was recorded as an acceptor photon with a specified probability (of 20%). Finally, background signal was generated for each channel as a Poisson process with the experimentally determined intensity.

## Results and Discussion

**TAR DNA Hairpins in the Absence of NC.** To study the effects of NC on the secondary structure and dynamics of the donor/acceptor labeled TAR DNA mutants, we begin with a simple, well-defined case: SMFRET of the -L3-L4 TAR DNA mutant in the absence of NC. Previously, we have shown that this mutant is well described by a fully closed hairpin structure with a mean FRET value of  $\langle E_A \rangle = 0.96$  and with undetectable  $E_A$  fluctuations.<sup>24</sup> A typical  $E_A(t)$  trajectory for -L3-L4 mutants is shown in Figure 1A. The durations of these trajectories are limited by photobleaching of Cy5.

One of the simplest ways to represent SMFRET data is in an ensemble single molecule  $E_A$  histogram, where ensemble refers to the combination of single molecule  $E_A$  histograms from many hairpins under the same conditions. To construct ensemble  $E_A$  histograms, the individual  $E_A(t)$  trajectories were boxcar time averaged, or “smoothed” with alternative bin times,  $\tau_B$ , of 3, 25, and 250 ms. In boxcar averaging a group of  $N$  adjacent  $E_A(t)$  points are averaged together where  $N = \tau_B/\tau_D$ . Here  $\tau_D$  is the time spacing in the original non-time-averaged data (i.e., 1 ms). As  $\tau_B$  is increased the signal-to-noise ratio of the data increases due to time averaging. FRET fluctuations are, however, also smoothed, or averaged, out of the data by this process, if they occur on a faster time scale than  $\tau_B$ . The boxcar time-averaged single molecule trajectories were combined to yield ensemble  $E_A$  histograms as shown in Figure 1B. The  $\tau_B = 3$  ms  $E_A$  histogram for this hairpin shows a single peak due to the C form (see Scheme 1). The width of the peak is primarily due to photon shot noise rather than  $E_A$  fluctuations, which is verified by  $E_A$  simulations (see below).

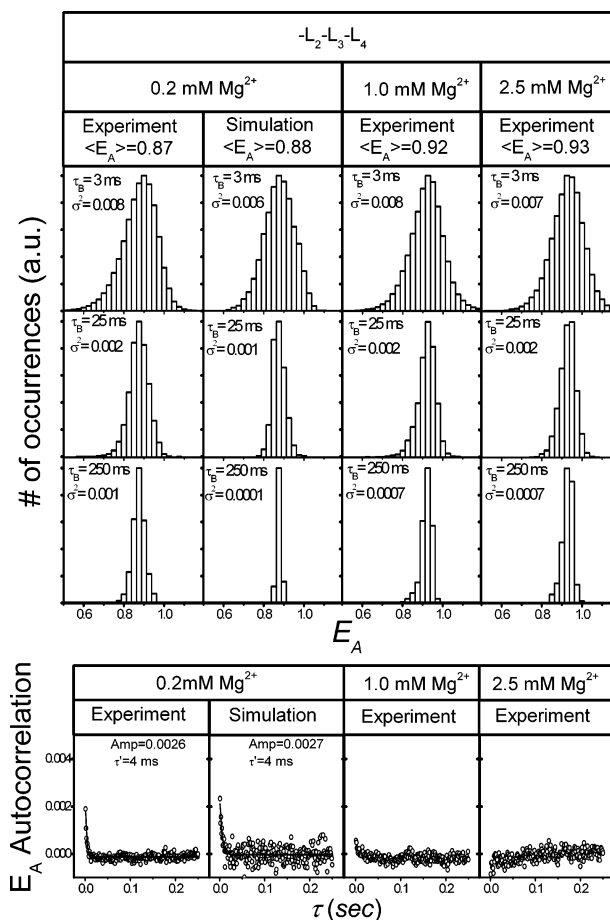
Simulated histograms, based on a single-state system with photon shot noise, are shown in Figure 1B (right panels) at each respective  $\tau_B$ , for comparison. The small differences in variance between the single state model and experimental values:

$$\Delta\sigma^2 = |\sigma_{\text{exp}}^2 - \sigma_{\text{theory}}^2| \quad (5)$$

reflect experimental error rather than actual  $E_A$  fluctuations. A careful comparison of individual trajectories suggests that small errors in establishing the background level in the donor and acceptor channels are the source of this error. Thus, for the 250 ms  $\tau_B$  histogram in Figure 1B the small residual error we observed,  $\sigma^2$ , will be denoted by  $\sigma_{\text{back}}^2$  to reflect the source of the error.

Ensemble  $E_A$  autocorrelation is another way to characterize  $E_A$  dynamics, where ensemble refers to the averaging of single molecule  $E_A$  autocorrelation curves from many hairpins under the same conditions. For TAR DNA hairpins in the presence of NC, the  $E_A$  autocorrelation exhibits exponential decays (see below) due to secondary structure fluctuations. For -L3-L4 in the absence of NC, no detectable component is observed within the experimental detection limit, i.e.,  $A_{AC} < 0.0005$  (Figure 1C, left panels). A system that exhibits  $A_{AC}$  below this detection limit can be considered a relatively “static” single state secondary structure. A simulated single molecule autocorrelation is also shown in Figure 1C (right panel). The undetectably small  $A_{AC}$  indicates that the -L3-L4 TAR DNA molecules in the absence of NC do not fluctuate in their structure. This static C form was observed for each TAR DNA mutant studied in the absence of NC at each concentration of  $\text{Mg}^{2+}$ .

**SMFRET of the -L2-L3-L4 TAR DNA/NC Complex.** Figure 2 shows NC-induced structural fluctuations for the -L2-L3-L4 TAR DNA mutant that only retains a single loop near the top of the stem (see Scheme 2). The shift in  $\langle E_A \rangle$  from the NC-free value of 0.96 (data not shown) to a value of 0.87 in the presence of 445 nM NC and 0.2 mM  $\text{Mg}^{2+}$  (Figure 2, column 1) is due to the presence of the open form at equilibrium.<sup>20</sup> Previous studies have suggested that a two-state equilibrium model can describe the fluctuations in short DNA hairpins.<sup>42,43</sup> In the current work, the experimental  $E_A$  histogram and autocorrelation data (Figure 2, column 1) are in near quantitative agreement with a two-state kinetic simulation (Figure 2, column 2), using a kinetic Monte Carlo (KMC) algorithm that is described in the Experimental Section and



**Figure 2.** (Top) Experimentally obtained ensemble  $E_A$  histograms for the -L2-L3-L4 TAR DNA mutant in the presence of 445 nM NC and three different  $\text{Mg}^{2+}$  concentrations. A simulated distribution is shown for the 0.2 mM  $\text{Mg}^{2+}$  data. (Bottom)  $E_A$  autocorrelation curves are shown for the same conditions.

elsewhere.<sup>41</sup> The small differences in variance  $\sigma^2$  between the simulation and experimental values in Figure 2 reflect experimental error  $\sigma_{\text{back}}^2$  rather than actual  $E_A$  fluctuations. As in the case of the NC free results (see Figure 1) the variance  $\sigma^2$  for all of the experimental  $\tau_B = 250$  ms  $E_A$  histograms in Figure 2 is not significantly larger than that expected from background error  $\sigma_{\text{back}}^2$ .

This implies that the opening and closing dynamics are both homogeneous and simple first-order within experimental error. Thus, the data do not indicate the presence of either long-lived intermediates or multiple opening/closing pathways for the -L2-L3-L4 TAR DNA/NC complex.

The required parameters for the simulation (Table 1) were determined by the following procedure. First, the experimental normalized  $E_A$  autocorrelation decay was fit by analytical results for a two-state interconversion as follows:

$$\frac{\langle E_A(t)E_A(t+\tau) \rangle}{\langle E_A(t) \rangle^2} - 1 = \frac{(E_{A,\text{open}} - E_{A,\text{close}})^2 k_{\text{open}} k_{\text{close}}}{(k_{\text{open}} E_{A,\text{open}} + k_{\text{close}} E_{A,\text{close}})^2} e^{-\lambda \tau} = \frac{(E_{A,\text{open}} - E_{A,\text{close}})^2 K_{\text{eq}}}{(K_{\text{eq}} E_{A,\text{open}} + E_{A,\text{close}})^2} e^{-\lambda \tau} \quad (6)$$

where  $\lambda$ , the rate constant, is simply  $\lambda = k_{\text{open}} + k_{\text{close}}$ .  $E_{\text{close}}$  was determined to be 0.96 from the experiment in the absence



**TABLE 1: Experimental SMFRET Parameters for the TAR DNA Mutants in the Presence of NC, and Associated Parameters Determined by Fitting the SMFRET Data, Using the Two-State Interconversion Model**

|                                 | -L3-L4                  | -L2-L3-L4/NC            |                         |                         | -L3-L4/NC               |                         |                         |
|---------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
|                                 | 0.2 mM Mg <sup>2+</sup> | 0.2 mM Mg <sup>2+</sup> | 1.0 mM Mg <sup>2+</sup> | 2.5 mM Mg <sup>2+</sup> | 0.2 mM Mg <sup>2+</sup> | 2.5 mM Mg <sup>2+</sup> | 5.0 mM Mg <sup>2+</sup> |
| $E_{A,close}$                   | 0.96                    | 0.96                    | 0.96                    | 0.96                    | 0.96                    | 0.96                    | 0.96                    |
| $\langle E_A \rangle$           | 0.96                    | 0.88                    | 0.92                    | 0.93                    | 0.77                    | 0.89                    | 0.92                    |
| $A_{AC}$                        |                         | 0.0026                  |                         |                         | 0.0070                  | 0.0032                  | 0.0012                  |
| rate (ms <sup>-1</sup> )        |                         | 0.24                    |                         |                         | 0.18                    | 0.084                   | 0.13                    |
| $K_{eq}$                        |                         | 5.23                    | 0.69                    | 0.46                    | 11.9                    | 2.34                    | 1.63                    |
| $E_{A,open}$                    |                         | 0.86                    | 0.86 <sup>a</sup>       | 0.86 <sup>a</sup>       | 0.75                    | 0.85                    | 0.90                    |
| $k_{open}$ (ms <sup>-1</sup> )  |                         | 0.20                    |                         |                         | 0.17                    | 0.059                   | 0.078                   |
| $k_{close}$ (ms <sup>-1</sup> ) |                         | 0.039                   |                         |                         | 0.014                   | 0.025                   | 0.048                   |

<sup>a</sup> The  $A_{AC}$  for these measurements was too small to fit, thus the -L2-L3-L4/NC/0.2 mM Mg<sup>2+</sup> values were used to determine  $K_{eq}$ . The inability to measure  $A_{AC}$  for these cases may be due to opening/closing rates that are too rapid to resolve by the experimental methods.

of NC. In the presence of NC,  $\langle E_A \rangle$  satisfies eq 7 for a two-state system:

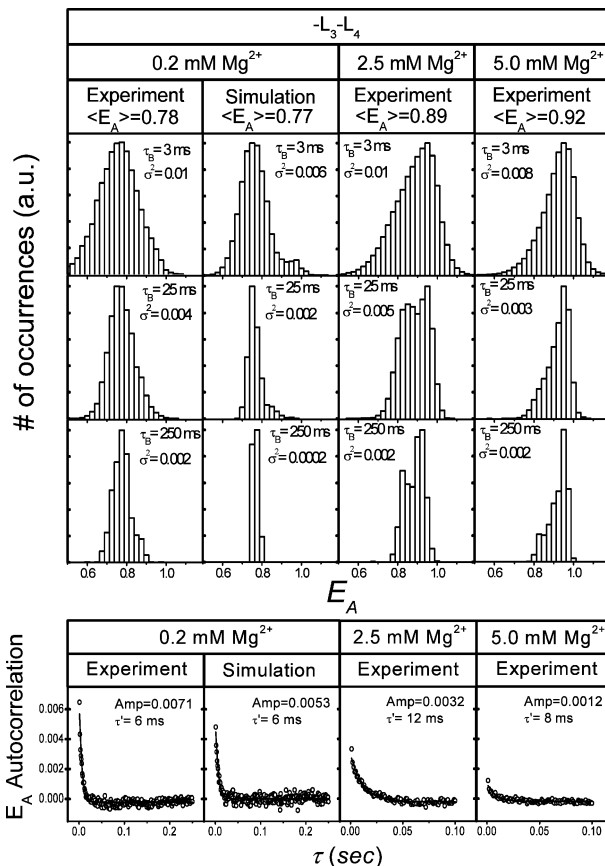
$$\langle E_A \rangle = \frac{K_{eq}E_{A,open} + 1 \cdot E_{A,close}}{1 + K_{eq}} \quad (7)$$

Equations 6 and 7 were used to calculate  $K_{eq}$  and  $E_{open}$  based on the experimentally obtained  $A_{AC}$  and  $\langle E_A \rangle$ . The rate constants,  $k_{open}$  and  $k_{close}$ , can thus be determined from fitting the  $E_A$  autocorrelation with a single decay constant,  $\lambda$ , and  $K_{eq}$  value.

Experimental data were also acquired for 1.0 (Figure 2, column 3) and 2.5 mM Mg<sup>2+</sup> (Figure 2, column 4), showing a shift in the equilibrium (see Table 1) toward the closed secondary structure as the Mg<sup>2+</sup> concentration is increased.

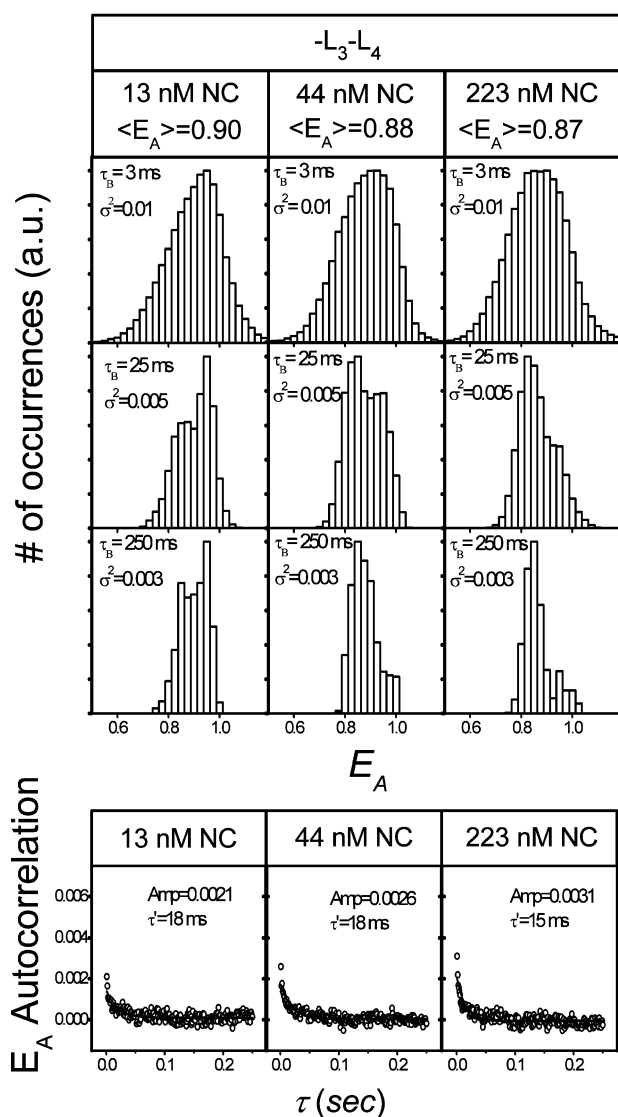
**SMFRET of -L3-L4 TAR DNA/NC Complex.**  $E_A$  data for NC induced dynamics of the -L3-L4 TAR DNA hairpin are shown in Figure 3. By analogy to the -L2-L3-L4 TAR DNA data, higher concentrations of Mg<sup>2+</sup> shift the equilibrium distribution toward the C form. However,  $\langle E_A \rangle$  for this mutant is significantly smaller than that observed for the -L2-L3-L4 mutant under similar conditions. This is consistent with opening through the L1-L2 regions. The experimental  $E_A$  histograms for different [Mg<sup>2+</sup>] and  $\tau_B$  are considerably broader than the best-fit two-state simulation of the data. The excess broadening in the experimental  $E_A$  data is even present at relatively long times as evidenced by the large  $\sigma^2$  in the  $\tau_B = 250$  ms histogram.

The presence in Figure 3 of broadening that significantly exceeds experimental error for the TAR DNA -L3-L4 mutant in the 250 ms  $\tau_B$  histogram indicates that the opening/closing process has a significant probability of occurring at times longer than 250 ms. This implies that the opening/closing is not a simple first-order process because the dominant relaxation time for the  $E_A$  fluctuations is on the 3–10 ms time scale. Qualitatively, the excess broadening indicates non-single-exponential end-to-end distance dynamics and, correspondingly, an underlying complexity in the secondary structure dynamics. For example, one or more long-lived intermediates along the reaction coordinate for the -L3-L4 mutant may exist, leading to two or more additional rate constants. Alternately, the complex secondary structure dynamics may be a reflection of parallel pathways for opening and closing kinetics, i.e., heterogeneous opening/closing kinetics. Hypothetical assignments for the nonexponential dynamics will be presented below. Furthermore, heterogeneity due to surface immobilization or chemical damage is assumed to be negligible in this system, consistent with the observations that analogous heterogeneity is observed regardless of the immobilization strategy and for various mutants and NC preparations. This is analogous to the heterogeneity that has been observed for several other nucleic acid/protein systems.<sup>25–27</sup>



**Figure 3.** (Top) Ensemble  $E_A$  histograms for the -L3-L4 TAR DNA mutant in the presence of 445 nM NC are presented for three different Mg<sup>2+</sup> concentrations. The simulated values at 0.2 mM Mg<sup>2+</sup> are included for comparison. (Bottom)  $E_A$  autocorrelation decays are compared for the three reaction conditions.

The trend in  $A_{AC}$  vs  $K_{eq}$  also demonstrates clear deviation from a simple two-state equilibrium model (Table 1). As shown previously for other TAR DNA constructs,<sup>20</sup> the effect of increasing [Mg<sup>2+</sup>] is to shift the equilibrium toward an equal distribution of C and Y forms, as demonstrated by the dependence of  $K_{eq}$  on [Mg<sup>2+</sup>] shown in Table 1. A simple two-state equilibrium model predicts that  $A_{AC}$  reaches a maximum value at an equal distribution of C and Y forms, which clearly deviates from the experimental results. In fact, the data shown in Table 1 and Figure 3 reveal that  $A_{AC}$  decreased when the equilibrium was shifted toward an equal distribution of C and Y forms. The experimental deviation from a simple two-state equilibrium model prediction for the dependence of  $K_{eq}$  on [Mg<sup>2+</sup>] trends is suggestive of more than one intermediate along the reaction coordinate. However, other explanations might be

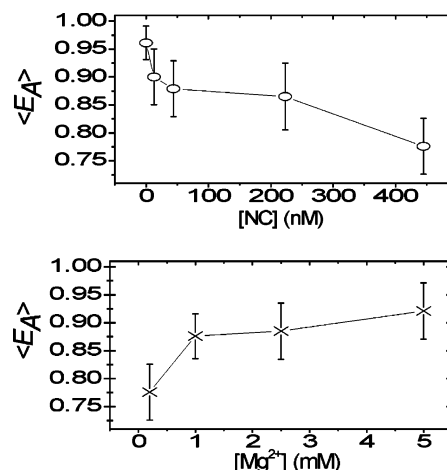


**Figure 4.** (Top) Ensemble  $E_A$  histograms for the -L3-L4 mutant at various NC concentrations are compared. (Bottom) The  $E_A$  autocorrelations for each ensemble are shown. All experiments were performed in the presence of 0.2 mM  $Mg^{2+}$  and 40 mM NaCl.

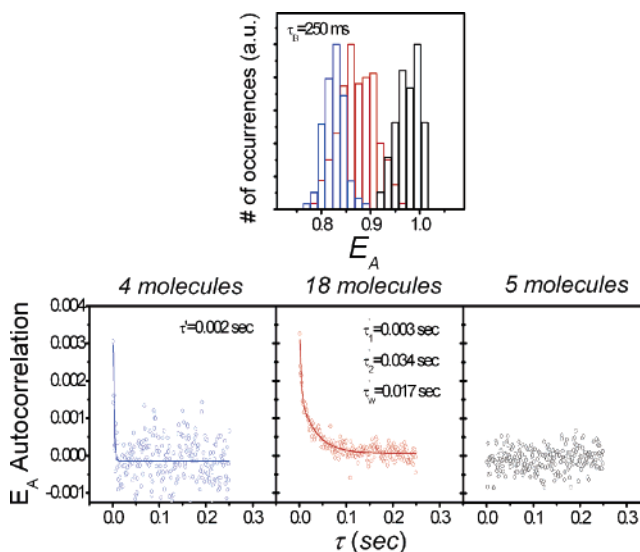
able to account for the observed deviation from a simple two-state model, such as a salt effect on the chain stiffness, which could alter the end-to-end distance and FRET values. Another indication that a simple two-state model does not adequately describe the opening/closing process of -L3-L4 TAR DNA in the presence of NC is the significant increase in the best-fit  $E_{A,open}$  values in Table 1 as a function of increasing  $Mg^{2+}$ . The apparent decrease in  $E_{A,open}$  as a function of decreasing  $Mg^{2+}$  apparently indicates that more than one open secondary structure of the hairpin is present at equilibrium. This furthermore implies that the experimental values obtained from the two-state analysis for -L3-L4 TAR DNA are qualitative at best since a three-state or even more complex treatment would be necessary to accurately describe this system.

Further studies of the NC's effects on -L3-L4 TAR DNA structural dynamics were performed as a function of NC concentration (Figure 4). In all cases, the -L3-L4 mutant exhibits heterogeneity on both short (5 ms) and long ( $>250$  ms) time scales.

As shown in Figures 3 and 4, the heterogeneous opening/closing dynamics in the -L3-L4 mutant were observed over a large range of NC and  $Mg^{2+}$  concentrations. The dependence



**Figure 5.** The trends in  $\langle E_A \rangle$  as a function of [NC] (top) and  $[Mg^{2+}]$  (bottom) are shown. The [NC] dependent experiments were performed at 0.2 mM  $Mg^{2+}$ , and the  $[Mg^{2+}]$  dependent experiments were performed at 445 nM NC. The lines are included as guides. The error bars denote  $\sigma$  at  $\tau_B = 250$  ms for each experiment.



**Figure 6.** (Top) The 250 ms  $\tau_B$  ensemble  $E_A$  histogram is shown for -L3-L4 TAR DNA, with 44 nM NC and 0.2 mM  $Mg^{2+}$ . The molecules have been sorted into primarily closed (black), primarily open (blue), and intermediate (red) structures, based on their single molecule  $E_A$  values. (Bottom) The corresponding subensemble  $E_A$  autocorrelation curves are shown, with the respective exponential curves and fitting parameters for each set. The  $A_{AC}$  curves for the intermediate molecules required double-exponential fitting. Each individual component, and the weighted time constant,  $\tau_w$ , are indicated in the plot.

of  $\langle E_A \rangle$  on the concentration of NC and  $Mg^{2+}$  is compared in Figure 5. The overall effects of NC and  $Mg^{2+}$  on the relative population of the Y form are in opposition to each other. Further experiments as a function of  $[Na^+]$  (data not shown) exhibited similar effects as that of  $Mg^{2+}$ , although higher  $[Na^+]$  were necessary to achieve comparable  $E_A$  values.

Further evidence that the opening/closing dynamics in the -L3-L4 mutant are heterogeneous is presented in Figure 6. Here the ensemble  $E_A$  histograms for the -L3-L4 mutant hairpins in the presence of 44 nM NC and 0.2 mM  $Mg^{2+}$  have been sorted and summed with respect to the single molecule  $E_A$  values. The resultant "subensemble"  $E_A$  histograms are plotted in color in the top panel in Figure 6. The corresponding subensemble autocorrelation curves are shown at the bottom of Figure 6. The black "subensemble" is a fully closed population with  $\langle E_A \rangle =$

0.98, while the red and blue subensembles are partially open structures with  $\langle E_A \rangle$  values of 0.87 and 0.83, respectively.

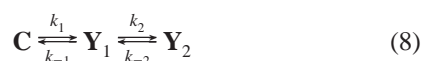
The subensemble of hairpins with the highest average  $E_A$  values (coded black in Figure 6) has FRET values that correspond to a fully closed structure, apparently the C form. The autocorrelation curve for this subensemble shows no measurable structural dynamics (i.e.,  $A_{AC} < 5 \times 10^{-4}$ ) indicating a relatively static hairpin structure with very little DNA end-to-end distance fluctuations. It is unlikely that this subensemble is due to hairpins with no bound NC, since at an NC concentration of 44 nM, approximately 2.5 NC molecules are estimated to be bound per hairpin assuming  $K_d \sim 100$  nM and 8 independent binding sites per hairpin. Furthermore, NC/hairpin collisions should occur on a shorter time scale than the  $>250$  ms lifetime of the subensemble of hairpins with the highest average  $E_A$  values, ensuring multiple NC binding/disassociation events per second.

In contrast, the  $E_A$  autocorrelation curves for the partially open “red” and “blue” subensembles show clear evidence of large-amplitude end-to-end distance fluctuation on a range of time scales from 2 to 17 ms. The “blue” subensemble exhibits the most rapid dynamics, on the  $\sim 2$  ms time scale near our instrumental resolution. The  $E_A$  autocorrelation times for the “red” subensemble are much longer and the decay is clearly nonsingle exponential.

It is important to emphasize that the basic assumption of this analysis is that the different subensembles are a consequence of kinetic heterogeneity on a  $>250$  ms time scale due to long-lived secondary structures and/or NC binding states. As described above, analogous heterogeneity has been observed for a broad range of NC and  $Mg^{2+}$  concentrations, e.g., the 250 ms histograms in Figure 4.

**Mechanistic Implications of the SMFRET DATA.** The observed SMFRET data suggest that the NC-induced destabilization in the L1-L2 regions of TAR DNA is in equilibrium between a relatively static structure with very little end-to-end distance fluctuations and a much more dynamic structure. Further, the structural fluctuations occur on multiple time scales, indicating a complex, multidimensional relationship between bound NC and the secondary structure in the L1-L2 regions of TAR DNA. Despite the extensive new observations, the available experimental data of various types on this process are still not sufficient to unequivocally determine the minimum number of kinetic species that are required to adequately describe this process. Perhaps, even more importantly, it is not possible for a given kinetic intermediate to specify the secondary structures, numbers of bound NC molecules, or the location of the bound NC molecules. Nevertheless, it is useful to consider simple kinetic models for the opening/closing process in an attempt to begin to identify possible physical origins of the observed effects.

The data in Figure 6 demonstrate that there is a significant population of long-lived closed hairpins, suggesting a kinetic model in which the initial opening event is rate limiting, as follows:

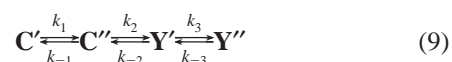


where  $k_1$ , the rate constant for initiating hairpin opening, may be as slow as 1/250 ms based on the SMFRET data. Within this model the millisecond time-scale dynamics observed in Figures 4 and 6 are assigned to secondary structure dynamics among the partially open Y forms. The relatively slow initiation of NC induced TAR DNA opening suggests that the transition

state for opening might involve thermally activated breaking of the Watson–Crick base pairs in the L1-L2 region, with only a partial lowering of the barrier for opening by the bound NC.

Interestingly, the model implied by eq 8 is in contrast with the two-state C/Y interconversion analysis of the -L2-L3-L4 TAR DNA data described above, since the latter analysis suggests a simple C/Y interconversion on the millisecond time scale. It is conceivable, however, that the model in eq 8 could also apply to the -L2-L3-L4 TAR DNA hairpin if both of the two states were assigned to different Y forms. This would require that very little C form be present at equilibrium for this hairpin in the presence of NC.

An alternative, equally plausible class of models that could account for the heterogeneous SMFRET dynamics is described by the following scheme:



in which there are two C forms that differ by secondary structure and/or NC binding number and configuration, leading to very different open rates. This model allows for both a long-lived C' form and short-lived C'' form. For example, C'' might be “activated” by an NC bound in a location, e.g., a bulge, that promotes the opening process. Thus, this model preserves the basic assignment of the SMFRET dynamics to the opening/closing process but still allows a subpopulation of static C forms.

At this point the data do not offer a means for distinguishing between the alternate kinetic models. Clearly more research will be necessary to determine whether either of these models captures the salient features of the NC induced opening/closing dynamics of TAR DNA hairpins, or whether even more elaborate kinetic treatments will be necessary.

## Conclusions and Summary

SMFRET was used to study NC-induced fluctuations of various DNA hairpins over a broader range of conditions and in more depth than in previous studies. As observed in other single molecule studies of nucleic acid structural fluctuations such as the hairpin ribozyme,<sup>44</sup> secondary structure dynamics over a broad distribution of time scales are observed.  $\tau_B$ -dependent histograms of  $E_A$  and autocorrelation analysis reveal complex dynamics that occur on two distinct time scales when both L1-L2 are present in the hairpin. In contrast, when L2 is absent, the fluctuations are constrained to a  $\sim 5$  ms lifetime. The experiments herein reveal a complex mechanism for secondary structure fluctuations, with dynamic processes occurring over a wide time range, i.e.,  $\sim 5$  to  $>250$  ms, and with the involvement of long-lived intermediates.

**Acknowledgment.** This work was supported by NIH grant GM65818 (P.B.), NIH grant GM65056 (K.M.-F.), NSF grant CHE-0347862 (D.E.M.), NIH postdoctoral National Research Service Award GM073534 (C.F.L.), and the Welch Foundation (P.B. and D.E.M.).

## References and Notes

- (1) Draper, D. E.; Reynaldo, L. P. *Nucleic Acids Res.* **1999**, *27*, 381–388.
- (2) Stockley, P. G. *Front. Mol. Biol.* **1999**, *21*, 22–58.
- (3) Doyle, M.; Jantsch, M. F. *J. Struct. Biol.* **2002**, *140*, 147–153.
- (4) Salas, M.; Freire, R.; Soengas, M. S.; Esteban, J. A.; Mendez, J.; Bravo, A.; Serrano, M.; Blasco, M. A.; Lazaro, J. M.; Blanco, L.; Gutierrez, C.; Hermoso, J. M. *FEMS Microbiol. Rev.* **1995**, *17*, 73–82.
- (5) Sevenich, F. W.; Langowski, J.; Weiss, V.; Rippe, K. *Nucleic Acids Res.* **1998**, *26*, 1373–1381.

- (6) Schultz, A.; Langowski, J.; Rippe, K. *J. Mol. Biol.* **2000**, *300*, 709–725.
- (7) Telesnitsky, A.; Goff, S. P. In *Retroviruses*; Coffin, J. M., Hughes, S. H., Varmus, H. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1997; pp 121–160.
- (8) Levin, J. G.; Guo, J.; Rouzina, I.; Musier-Forsyth, K. *Prog. Nucleic Acid Res. Mol. Biol.* **2005**, *80*, 217–286.
- (9) Zhang, Y.; Barklis, E. *J. Virol.* **1995**, *69*, 5716–5722.
- (10) Zhang, Y.; Qian, H.; Love, Z.; Barklis, E. *J. Virol.* **1998**, *72*, 1782.
- (11) Feng, Y. X.; Campbell, S.; Harvin, D.; Ehresmann, B.; Ehresmann, C.; Rein, A. *J. Virol.* **1999**, *73*, 4251–4256.
- (12) Chan, B.; Weidemaier, K.; Yip, W.-T.; Barbara, P. F.; Musier-Forsyth, K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 459–464.
- (13) Wu, T.; Guo, J.; Bess, J.; Henderson, L. E.; Levin, J. G. *J. Virol.* **1999**, *73*, 4794–4805.
- (14) Guo, J.; Wu, T.; Anderson, J.; Kane, B. F.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Levin, J. G. *J. Virol.* **2000**, *74*, 8980–8988.
- (15) You, J. C.; McHenry, C. S. *J. Biol. Chem.* **1994**, *269*, 31491.
- (16) Rein, A.; Henderson, L. E.; Levin, J. G. *Trends Biochem. Sci.* **1998**, *23*, 297–301.
- (17) Johnson, P. E.; Turner, R. B.; Wu, Z. R.; Hairston, L.; Guo, J.; Levin, J. G.; Summers, M. F. *Biochemistry* **2000**, *39*, 9084–9091.
- (18) Azoulay, J.; Clamme, J. P.; Darlix, J. L.; Roques, B. P.; Mély, Y. *J. Mol. Biol.* **2003**, *326*, 691–700.
- (19) Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J. P.; Ficheux, D.; Roques, B.; Darlix, J. L.; Mély, Y. *J. Mol. Biol.* **2003**, *328*, 95–108.
- (20) Cosa, G.; Harbron, E. J.; Zeng, Y.; Liu, H. W.; O'Connor, D. B.; Eta-Hosokawa, C.; Musier-Forsyth, K.; Barbara, P. F. *Biophys. J.* **2004**, *87*, 2759–2767.
- (21) Urbaneja, M. A.; Wu, M.; Casas-Finet, J. R.; Karpel, R. L. *J. Mol. Biol.* **2002**, *318*, 749–764.
- (22) Williams, M. C.; Rouzina, I.; Bloomfield, V. A. *Acc. Chem. Res.* **2002**, *35*, 159–166.
- (23) Lu, H. P.; Iakoucheva, L. M.; Ackerman, E. J. *J. Am. Chem. Soc.* **2001**, *123*, 9184–9185.
- (24) Zhuang, X.; Bartley, L. E.; Babcock, H. P.; Russell, R.; Ha, T.; Herschlag, D.; Chu, S. *Science* **2000**, *288*, 2048.
- (25) Rothwell, P. J.; Berger, S.; Kensch, O.; Felekyan, S.; Antonik, M.; Wohrl, B. M.; Restle, T.; Goody, R. S.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1655.
- (26) Ha, T. *Biochemistry* **2004**, *43*, 4055–4063.
- (27) Ha, T.; Rasnik, I.; Cheng, W.; Babcock, H. P.; Gauss, G.; Lohman, T. M.; Chu, S. *Nature* **2002**, *419*, 638–641.
- (28) Xie, Z.; Srividya, N.; Sosnick, T. R.; Pan, T.; Scherer, N. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 534–539.
- (29) Tan, E.; Wilson, T. J.; Nahas, M. K.; Clegg, R. M.; Lilley, D. M. J.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9302.
- (30) Lu, P. H.; Xun, L.; Xie, X. S. *Science* **1998**, *282*, 1877.
- (31) Zhuang, X.; Rief, M. *Curr. Opin. Struct. Biol.* **2003**, *13*, 88–97.
- (32) Onoa, B.; Tinoco, I., Jr. *Curr. Opin. Struct. Biol.* **2004**, *14*, 374–379.
- (33) Onoa, B.; Dumont, S.; Liphardt, J.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2003**, *299*, 1892–1896.
- (34) Previous work, using fluorescence correlation spectroscopy to study a similar cTAR DNA and truncated NC system, reported conformational fluctuations on the microsecond time scale (see refs 18 and 19). This technique, because it is limited by the diffusion-controlled sampling lifetime, is not capable of measuring slow dynamics, such as those reported here, that occur on millisecond (and much longer) time scales.
- (35) Liu, H. W.; Cosa, G.; Landes, C. F.; Zeng, Y.; Mullen, D. G.; Barany, G.; Musier-Forsyth, K.; Barbara, P. F. *Biophys. J.* **2005**, *89*, 3470–3479.
- (36) Lee, B. M.; De Guzman, R. N.; Turner, B. G.; Tjandra, N.; Summers, M. F. *J. Mol. Biol.* **1998**, *279*, 633–649.
- (37) Zuker, M. *Nucleic Acids. Res.* **2003**, *31*, 3406–3415.
- (38) Zuker, M.; Mathews, D. H.; Turner, D. H. *Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999.
- (39) Ha, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 287–292.
- (40) Harada, Y.; Sakurada, K.; Aoki, T.; Thomas, D. D.; Yanagida, T. *J. Mol. Biol.* **1990**, *216*, 49–68.
- (41) Wang, Z.; Makarov, D. E. *J. Phys. Chem. B* **2003**, *107*, 5617–5622.
- (42) Bonnet, G.; Krichevsky, O.; Libchaber, A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8602–8606.
- (43) Ritort, F.; Bustamante, C.; Tinoco, I., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13544–13548.
- (44) Rueda, D.; Bokinsky, G.; Rhodes, M. M.; Rust, M. J.; Zhuang, X.; Walter, N. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10066–10071.