JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Tertiary DNA Structure in the Single-Stranded hTERT Promoter Fragment Unfolds and Refolds by Parallel Pathways via Cooperative or Sequential Events

Zhongbo Yu,[†] Vanessa Gaerig,[‡] Yunxi Cui,[†] HyunJin Kang,[‡] Vijay Gokhale,[‡] Yuan Zhao,[†] Laurence H. Hurley,*[,‡,§,∥] and Hanbin Mao*[,†]

[†]Department of Chemistry and Biochemistry and School of Biomedical Sciences, Kent State University, Kent, Ohio 44242, United States

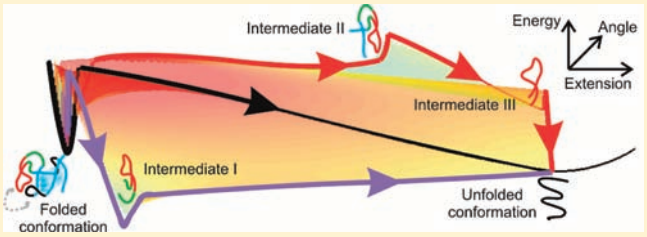
[‡]College of Pharmacy, University of Arizona, 1703 East Mabel Street, Tucson, Arizona 85721, United States

[§]Arizona Cancer Center, 1515 North Campbell Avenue, Tucson, Arizona 85724, United States

[∥]BIO5 Institute, 1657 East Helen Street, Tucson, Arizona 85721, United States

Supporting Information

ABSTRACT: The discovery of G-quadruplexes and other DNA secondary elements has increased the structural diversity of DNA well beyond the ubiquitous double helix. However, it remains to be determined whether tertiary interactions can take place in a DNA complex that contains more than one secondary structure. Using a new data analysis strategy that exploits the hysteresis region between the mechanical unfolding and refolding traces obtained by a laser-tweezers instrument, we now provide the first convincing kinetic and thermodynamic evidence that a higher order interaction takes place between a hairpin and a G-quadruplex in a single-stranded DNA fragment that is found in the promoter region of human telomerase. During the hierarchical unfolding or refolding of the DNA complex, a 15-nucleotide hairpin serves as a common species among three intermediates. Moreover, either a mutant that prevents this hairpin formation or the addition of a DNA fragment complementary to the hairpin destroys the cooperative kinetic events by removing the tertiary interaction mediated by the hairpin. The coexistence of the sequential and the cooperative refolding events provides direct evidence for a unifying kinetic partition mechanism previously observed only in large proteins and complex RNA structures. Not only does this result rationalize the current controversial observations for the long-range interaction in complex single-stranded DNA structures, but also this unexpected complexity in a promoter element provides additional justification for the biological function of these structures in cells.



INTRODUCTION

Protein and RNA are well-known to assume tertiary structures with biological functions. While single-domain proteins show cooperative, two-state folding and unfolding processes,[1] intermediates are more common in the folding and unfolding of complex proteins[2,3] and RNA structures.[4,5] In RNA, local interactions between secondary elements are responsible for the hierarchical nature of the structure.[4−6] Until recently, in complete contrast to protein and RNA, it was believed that DNA serves solely as the repository for genetic information and does not have direct biological functions. Recent insight into non-B-DNA structures, such as G-quadruplexes,[7] is challenging this stereotypical profile of DNA. There is now compelling evidence that G-quadruplexes exist naturally in cells,[8−10] and they can serve as transcription regulators for various genes, including c-MYC[8] and c-KIT.[11]

G-Quadruplexes, which consist of four runs of two or more contiguous guanines connected through G-tetrads,[12] are found to be abundant in promoter elements throughout the human genome.[13] Although still controversial,[14] in vitro evidence points to the existence of high-order interactions between neighboring G-quadruplexes[15,16] in single-stranded DNA (ss) sequences, such as those of telomeres, which can host multiple G-quadruplexes. This suggests that, like RNA or protein, tertiary DNA structures may also form. Despite their significant structural differences, protein and RNA share similar energy minima in the folding energy landscape representative of a large molecule.[17] This leads to alternative folding pathways for both species, a process known as the kinetic partition mechanism.[17] It would be interesting to see whether this unifying mechanism can apply to complex DNA structures as well.

Using a mechanical unfolding and refolding approach in a home-built laser-tweezers instrument,[18,19] we have observed a complex structure in an ssDNA fragment from the core promoter of human telomerase reverse transcriptase (hTERT).

Received: November 4, 2011
Published: February 28, 2012

   dx.doi.org/10.1021/ja210399h | J. Am. Chem. Soc. 2012, 134, 5157−5164

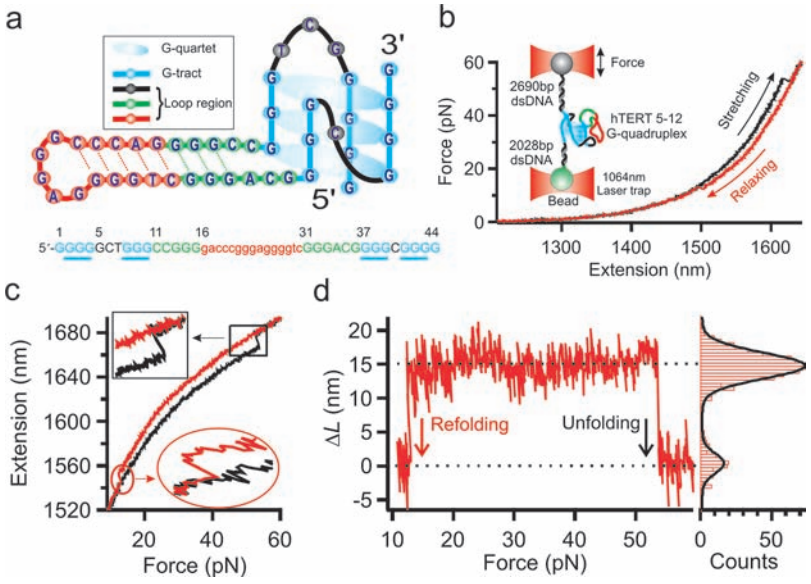


Figure 1. Cooperative unfolding and refolding in the hTERT 5−12 fragment. (a) The structure formed in the hTERT 5−12 fragment. (b) Force−extension curves of a DNA construct that contains the hTERT fragment. Black and red data points (100 Hz bandwidth) represent stretching and relaxing processes, respectively. The inset shows the experimental setup, in which a single DNA construct is tethered between two beads trapped by laser tweezers. (c) Extension−force curves show cooperative unfolding (black) and refolding (red) events. (d) Difference in the extension, Δx, between the unfolding and refolding traces in (c) is converted to the change-in-contour-length (ΔL) between 10 and 60 pN (the ΔL−F plot). A histogram of ΔL from the ΔL−F plot is shown to the right. The black curve represents a two-peak Gaussian fitting.

This structure is consistent with that determined by DMS footprinting and circular dichroism (CD),[20] which reveals a G-quadruplex with a long hairpin-containing loop. We have shown here that unfolding/refolding events of this structure follow either cooperative or sequential pathways. This observation directly supports the kinetic partition mechanism previously seen only in the folding of complex protein[2,3] and RNA structures.[4,5] Among the three intermediates we have observed for the sequential unfolding/refolding pathways, the 15-nucleotide (nt) hairpin is the critical species that serves as a nucleus in the refolding. Removal of this hairpin by mutation or hybridizing with a complementary strand dramatically reduces the cooperative unfolding and entirely abrogates the cooperative refolding events. Comparison of the change in free energy of unfolding (ΔG_{unfold}) between the sequential and cooperative pathways reveals a tertiary interaction (~6 kcal/mol) between the G-quadruplex and the hairpin. The partition of the DNA complex into a conformation with or without this tertiary interaction may result in cooperative or sequential unfolding/refolding pathways. This helps to explain the current controversial observations on the existence of DNA tertiary interactions.[14−16]

RESULTS AND DISCUSSION

The hTERT 5−12 Structure Follows Either Cooperative or Sequential Unfolding/Refolding Pathways. To investigate the unfolding and refolding processes of a complex DNA structure with potential tertiary interactions, we focused on a DNA fragment (Figure 1a), 5′-GGG GGC TGG GCC GGG gac ccg gga ggg gtc GGG ACG GGG CGG GG, from the core promoter (−90 to −46, or hTERT 5−12) of hTERT. We sandwiched this ssDNA fragment between 2690- and 2028-bp double-stranded DNA (dsDNA) handles, which were immobilized to the two optically trapped beads (Figure 1b, inset). A force−ramp assay was carried out at the force-loading rate of 5.5 pN/s in 10 mM Tris buffer with 100 mM K^+ (pH 7.4) to collect force−extension (F−X) curves (see Experimental Section). To analyze unfolding or refolding transitions that are represented by a sudden change in the tension or extension in the F−X curves (Figure 1b), we switched the F−X coordinates (Figure 1c), which were then transformed to a plot of change-in-contour-length (ΔL) versus force, or ΔL−F plot (Figure 1d, left), using a worm-like-chain model (see Experimental Section). The histogram derived from each trace allows clear identification of a transition process (ΔL = 15 nm population) from background (ΔL = 0 nm population) (Figure 1d, right). The ΔL determined in this manner matches remarkably well with that averaged from single F−X curve measurements (see Experimental Section and Figure S1). After 2448 plots were analyzed, two types of transitions were observed. One was a large, single-step transition with an average ΔL of 16 ± 2 nm (Figures 1d and 2), which represents cooperative unfolding or refolding; the other was composed of a series of smaller steps, which indicates sequential kinetic pathways (Figure 2a). As was clearly demonstrated here, this novel approach allows straightforward and intuitive determination of small transitions that are often obscured in conventional F−X curves (Figure S1). The ΔL of the cooperative transitions was consistent with unfolding of the 42-nt G-quadruplex−hairpin hybrid previously determined (Figure 1a, see Experimental Section for calculation).[20]

Population analysis revealed a significant portion (28% in unfolding and 17% in refolding, Figure 2f) of the events undergoing cooperative transitions. These events suggest the existence of a higher order interaction in the complex DNA structure,[5,21] likely between the G-quadruplex and the hairpin. With the current unfolding geometry (Figure 1b), the G-quadruplex should be ruptured first, which would simultaneously destroy the tertiary interaction between the G-quadruplex and the hairpin, leading to the immediate unfolding of the hairpin whose rupture force is lower than that for the G-quadruplex.[19,20,22]

dx.doi.org/10.1021/ja210399h | J. Am. Chem. Soc. 2012, 134, 5157−5164

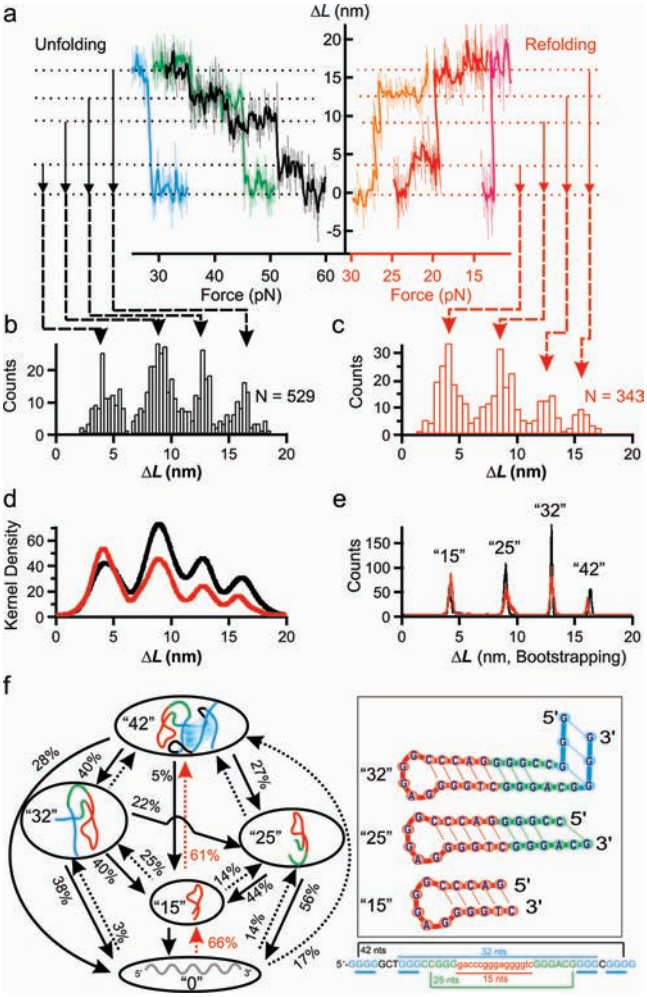


Figure 2. Three intermediates in the sequential unfolding/refolding transitions. The ΔL–F plots for typical unfolding and refolding transitions (a). Data with 100 Hz (thin traces) and 10 Hz (thick traces) bandwidths are plotted together. Population histograms for fully folded ("42"); 32-nt ("32"), 25-nt ("25"), and 15-nt ("15") intermediates; and fully unfolded ("0") species are shown in (b) and (c). (d) Kernel density estimation of the unfolding (black) and refolding (red) processes. (e) Histograms (black for the unfolding and red for the refolding) of the population based on bootstrapping analyses. (f) Schematic of unfolding (black solid arrows) or refolding (dotted arrows, major pathways are highlighted with red arrows) transitions for structures (shown to the right) in the hTERT 5–12 fragment. Percentage value depicts the probability for a particular transition.

Two possible scenarios may explain the 17% cooperative refolding. First, the G-quadruplex folded before the hairpin. Limited by the refolding trajectory (Figure 1b), we cannot follow subsequent formation kinetics of the hairpin. However, because the folding of the G-quadruplex is slower than that for the hairpin,[19,22] together with the direct measurement of the two refolding rate constants (shown below), such a refolding order is unlikely. In the second scenario, the hairpin and G-quadruplex refolded simultaneously. This may occur as the hairpin facilitates the preassembly of the four G-rich tracts into a G-quadruplex. When these experiments were repeated in a 10 mM Tris buffer (pH 7.4) with 100 mM Li^+, the 16 nm (or the 42-nt) population disappeared (Figure S2), which confirmed that the presence of a G-quadruplex is required for tertiary structure formation in the DNA complex, as Li^+ is known to inhibit quadruplex formation.

The 15-nt Hairpin Is the Key Intermediate in the Unfolding and Refolding Pathways. The majority of unfolding (72%) or refolding (83%) events were sequential. These observations strongly suggest that the complex DNA structure is hierarchical, which has been observed previously only in complex proteins and RNAs.[4,5] To reveal the number of species in sequential processes, we measured ΔL for the transition between each folded structure and the unfolded ssDNA. The ΔL histogram reveals four populations for both unfolding (Figure 2b, N = 529) and refolding (Figure 2c, N = 343) processes. The four populations were confirmed by the probability density estimation for all transitions based on a Gaussian kernel calculation[23] (Figure 2d, see Experimental Section). To accurately determine the ΔL for each species, we then used a bootstrapping analysis to identify the most likely peaks from the kernel density estimation (see Experimental Section). Figure 2e shows that the size of each structure exactly matches both the unfolding and the refolding trajectories, suggesting that the same species are involved in both processes. After converting ΔL into the number of nucleotides (nt) (see Experimental Section), a fully folded conformation with 42-nt and three intermediates that contain 32-, 25-, and 15-nt was clearly identified (Figure 2e). The 5′-end to 3′-end unfolding geometry (Figure 1b) led us to propose intermediate structures in Figure 2f. While they all resemble hairpins, the 32-nt species has the longest stem, in which the last three Hoogsteen guanine pairs are residues from the G-quadruplex. The 25-nt species contains a stem with two bulged nucleotides (see Figure S3a for a stabilized conformation from simulation), whereas the 15-nt species has the canonical hairpin structure with a pentaloop and a stem comprising five Watson–Crick base pairs.

Figure 2f summarizes all of the observed transitions among the three intermediates ("32", "25", and "15"), the fully folded structure ("42"), and the unfolded ssDNA ("0") with the probability for each process. For clarity, Figure 3a summarizes the major unfolding/refolding events with a probability >25%.

The fully folded G-quadruplex–hairpin hybrid undergoes three major unfolding trajectories: the cooperative unfolding (28%), through the 32-nt species (40%), or through the 25-nt intermediate (27%). Interestingly, the same 15-nt hairpin appeared in the latter two routes. Calculation revealed that starting with the fully folded structure, ~28% of all unfolding trajectories encountered the 15-nt hairpin intermediate. During refolding, the importance of the 15-nt hairpin was more obvious (Figure 3a): 66% of the population folded through this species. In contrast, only 14% of the population folded back to the 25-nt intermediate, while 17% underwent cooperative refolding. Once the 15-nt hairpin was formed, the majority of the hairpin (61%) directly folded to the G-quadruplex–hairpin hybrid. The coexistence of sequential and cooperative refolding pathways directly supports the kinetic partition mechanism previously associated only with complex proteins or RNA structures.[17]

Kinetics and Thermodynamic Analyses Reveal a Tertiary Interaction between the Hairpin and the G-Quadruplex. The critical role of the 15-nt hairpin was again manifested in the kinetic analyses, in which only major transitions (>25% probability) were processed for accurate determination of rate constants (Figure 3a). For a representative cooperative unfolding event ("42"→"0"), the Dudko model[24] (Figure 3b) and the Evans model[25] (Figure S4)

dx.doi.org/10.1021/ja210399h | J. Am. Chem. Soc. 2012, 134, 5157–5164

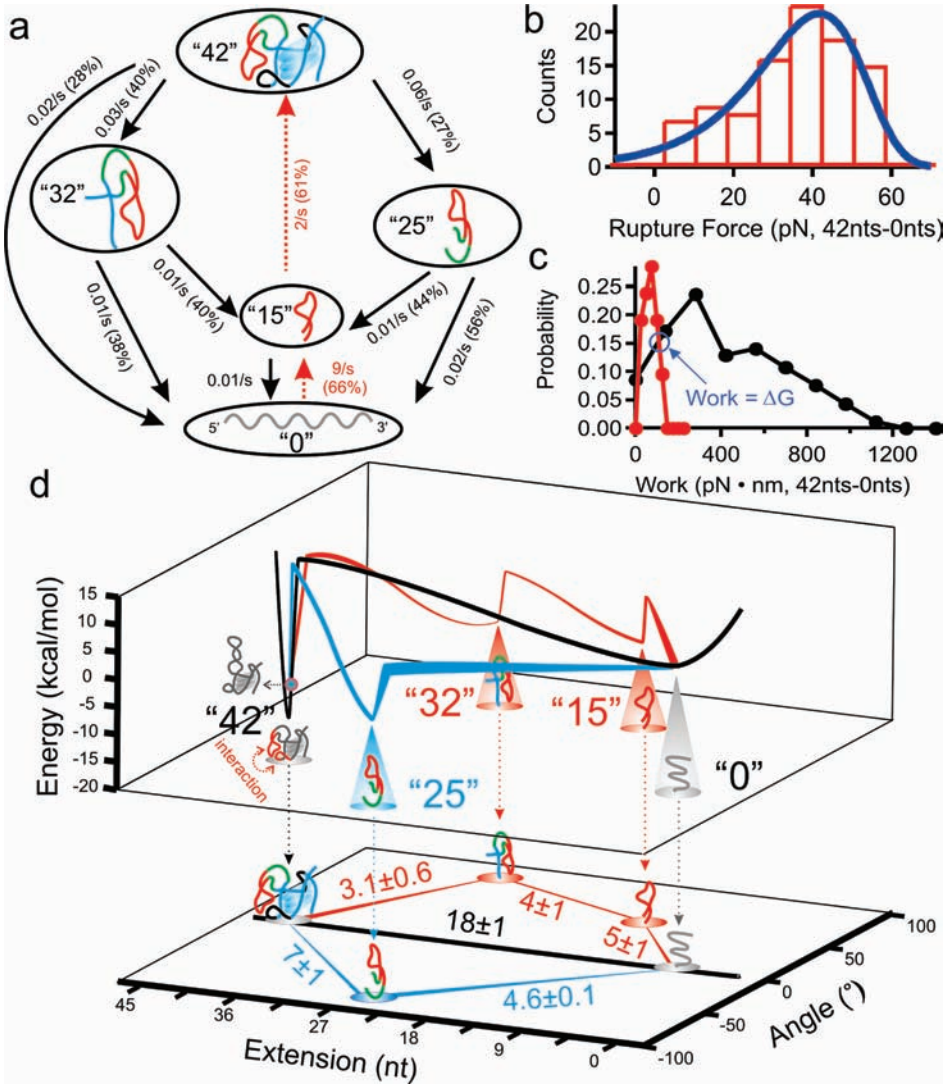


Figure 3. Kinetics and thermodynamics of the major transitions. (a) Unfolding (black solid arrows) or refolding (red dotted arrows) transitions (>25% probability) labeled with rate constants. (b) Histogram of the unfolding force for the "42"→"0" transition. Blue curve is the Dudko fitting.[24] (c) Histogram of the work associated with the unfolding (black) or refolding (red) of the "42"→"0" transition. The cross point is the change in free energy (ΔG) for the transition.[27,28] (d) Free energy versus extension (nt) and angle (deg) (see Experimental Section) for "42"→"0" (black), "42"→"25"→"0" (blue), and "42"→"32"→"15"→"0" (red) trajectories. The colored circle above "42" depicts the folded structure without tertiary interactions. Bottom projection shows ΔG (mean ± s.d., kcal/mol) between different species.

provided identical k_{unfold} values of 0.02 s^{-1}. In addition, the Dudko model revealed an activation energy, $\Delta G^{\ddagger}$ = 29 kcal/mol, for this event (Figure 3b). These values were then used to calculate the unfolding activation energy for other processes (see Experimental Section and Table 1).

Table 1. Thermodynamics, Kinetics, and Unfolding/Refolding Angles of the Major Transitions in the hTERT 5−12 Fragment without the 15-nt Hairpin Complement at 23 °C

transition	$x^{\ddagger}$ (nm)	ΔG (kcal/mol)	$\Delta G^{\ddagger}$ (kcal/mol)	angle (deg)
42−0	0.33 ± 0.01	18 ± 1	29 ± 2	0.0
42−32	0.33 ± 0.02	3.1 ± 0.6	29 ± 3	86.7
42−25	0.23 ± 0.02	7 ± 1	28 ± 3	−80.0
32−15	0.38 ± 0.02	4 ± 1	10 ± 1	−82.0
25−0	0.37 ± 0.03	4.6 ± 0.1	9 ± 2	78.9
15−0	0.41 ± 0.02	5 ± 1	10 ± 1	−76.3

Similar refolding analyses indicate that the 15-nt hairpin has the fastest refolding kinetics (k = 9 s^{-1}, Figure 3a), suggesting a seeding role of this hairpin to the fully folded G-quadruplex−hairpin hybrid. Although a long loop discourages the formation of a G-quadruplex,[26] the formation of the 15-nt hairpin in the loop apparently remediates this effect by gathering distal G-rich tracts for the G-quadruplex assembly. The fact that the hairpin folds much faster than the G-quadruplex (9 vs 2 s^{-1}, Figure 3a) is consistent with the literature[19,22] and supports our argument that a tertiary interaction between the G-quadruplex and the hairpin is responsible for cooperative transitions (see above).

The tertiary interaction was further analyzed from a perspective of change in free energy (ΔG) using the Crooks theorem[27,28] (see Figure 3c for the "42"→"0" transition; for the remainder, see Figure S5). $\Delta G_{42\rightarrow 0}$ for the cooperative process (18 kcal/mol, the black process in Figure 3d) was significantly higher than those for sequential pathways (12.1 and 11.6 kcal/mol for "42"→"32"→"15"→"0" (red) and "42"→"25"→"0" (blue) trajectories, respectively, Figure 3d). This result suggests

dx.doi.org/10.1021/ja210399h | J. Am. Chem. Soc. 2012, 134, 5157−5164

that folded structures can partition into a conformation with or without a tertiary interaction, which is on the order of 6 kcal/mol. We surmise that the long-range interaction can result from the docking of the hairpin to the G-quadruplex, which may be aided by the base-pairing between the hairpin loop and one of the G-quadruplex loops. Figure 3d summarizes the 3D free energy landscape versus distance (nt) and angle (deg) for major trajectories (see Table 1 and Figure S6 for details).

Tertiary Interaction Is Perturbed by a DNA Fragment Complementary to the 15-nt Hairpin. The seeding role of the 15-nt hairpin in the refolding processes was further confirmed in the experiments in which 10 μM ssDNA complementary to the 15-nt hairpin was incubated with 0.2 nM hTERT 5−12 fragment (see Experimental Section). Footprinting has shown no significant change in the DNA structure in the presence of the ssDNA complement (Figure 4a). However, the probability for pathways involving the 15-nt hairpin was much reduced (Figure 4b and Figure S7). In contrast, the 32-nt and 25-nt species became predominant during unfolding. During the refolding processes that became less likely to occur (Figure S7), the 25-nt species evolved to be the predominant species with the fastest rate (Figures 4b and S8). These results were expected as the 15-nt hairpin was disturbed after hybridizing with the ssDNA. The fact that the 25-nt intermediate became predominant during refolding further supports the "seeding" role of the 15-nt hairpin, without which the 25-nt species would function as an alternative to bring together distal G-rich tracts for the quadruplex assembly. Computer simulation confirmed that the 25-nt species is stable in the presence of the ssDNA complement (Figure S3b).

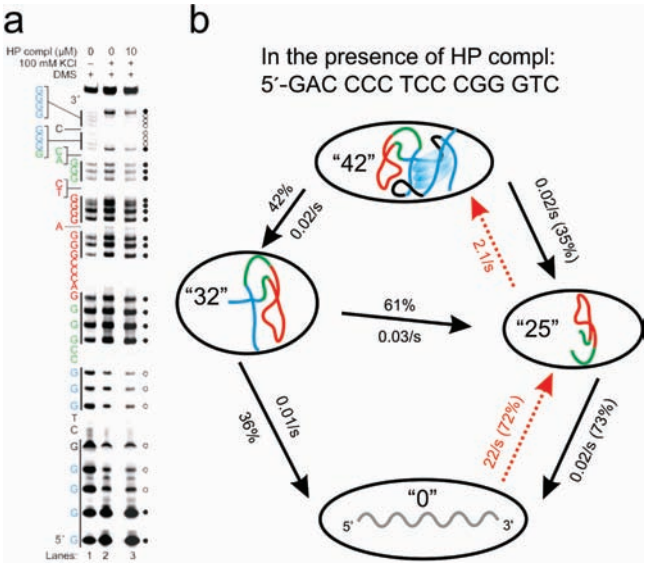


Figure 4. Effect of the 15-nt hairpin complement (HP compl) on the hTERT 5−12 fragment. (a) DMS footprinting of the hTERT fragment without (lane 2) and with (lane 3) HP compl shows that the complement has no effect on the footprinting pattern. Circles to the right of the gel indicate protection (○) and cleavage (●) by DMS. Vertical black lines to the left indicate G-tracts. (b) Major transitions (>25% probability) of the hTERT fragment in the presence of the HP compl (unfolding events are shown with black solid arrows; refolding events are shown with red dotted arrows).

The addition of the ssDNA fragment obliterated the cooperative refolding and dramatically reduced the cooperative unfolding events (Figures 4b and S7). This strongly supports that the tertiary interaction between the G-quadruplex and the hairpin is responsible for the cooperative transitions. With the loss of the 15-nt hairpin structure, the premise for the long-range interaction no longer exists, which leads to the drastic reduction of the cooperative unfolding or refolding processes. The observation that tertiary interaction is responsible for the cooperative unfolding/refolding transitions is in full agreement with previous findings.[29−35]

Further support for the critical role of the 15-nt hairpin came from the unfolding and refolding of a mutant, 5′-GGG GGC TGG GCC GGG gac ccg gga g**aa** **act** GGG ACG GGG CGG GG (Figure S9). The bold, underlined sequence in this fragment is the mutation to prevent the formation of the 15-nt hairpin, which was confirmed by the laser-tweezers experiments (Figures S9a and b). In addition, only 4% of fully folded structure was observed among all populations, and the cooperative unfolding constituted 4% of all events (cooperative refolding was not observed). These results corroborated the previous finding that the 15-nt hairpin plays a pivotal role for the tertiary interaction in the DNA complex.

■ CONCLUSIONS

In summary, the mechanical unfolding and refolding of a complex DNA structure in the single-stranded hTERT promoter fragment suggest a tertiary interaction between the G-quadruplex and the hairpin in the structure. Among the three intermediates identified in the sequential or cooperative unfolding/refolding pathways, the 15-nt hairpin demonstrates its indispensable role in these processes. That this structure shows both cooperative and hierarchical unfolding/refolding pathways not only supports the unifying kinetic parallel mechanism previously seen only in RNA and complex proteins, it also resolves current contradictory observations on the presence of higher order DNA interactions, as folded structures can partition into a conformation with or without tertiary interactions. While this study does not take into account other proximal interactions that might occur with an adjacent G-quadruplex in the hTERT promoter,[20] our results provide fundamental structural and kinetic insights into complex DNA structures with tertiary interactions in a single-stranded DNA context in vitro. Because telomerase is a critical enzyme involved in cancer and senescence,[36] we anticipate the higher order interaction and the intermediates revealed here to be instrumental in the development of drugs for cancer or other age-related diseases by targeting the hTERT promoter regions.

■ EXPERIMENTAL SECTION

DNA Engineering. The DNA construct used in the single-molecule assay consists of a single-stranded G-quadruplex-forming sequence in the promoter region of human telomerase (5′-GGG GGC TGG GCC GGG gac ccg gga ggg gtc GGG ACG GGG CGG GG, hTERT 5−12) sandwiched between two dsDNA handles. To reduce the interference from dsDNA handles on the G-quadruplex-forming sequence, five deoxythymidine residues were introduced at each end of the hTERT 5−12 fragment. The DNA construct was prepared according to procedures described previously.[37] Briefly, an ssDNA/dsDNA hybrid with two restriction enzyme sites, *Eag*I and *Xba*I, was ligated to two dsDNA handles. One of the dsDNA handles (2690 bp) was prepared by *Eag*I and *Sac*I (New England Biolabs) digestions of a pEGFP vector (Clontech, Mountain View, CA) followed by agarose gel purification. The digoxigenin was then introduced at the *Sac*I end by terminal deoxynucleotidyl transferase (New England Biolabs). The other dsDNA handle was a 2028 bp PCR product amplified from the

dx.doi.org/10.1021/ja210399h | J. Am. Chem. Soc. 2012, 134, 5157−5164

pBR322 plasmid (New England Biolabs). This handle has a biotin at one end and an *Xba*I site at the other. Final ligation between the hTERT 5−12 containing the dsDNA−ssDNA hybrid and the two dsDNA handles was achieved by T4 DNA ligase (New England Biolabs).

Single-Molecule Force−Ramp Assay. To immobilize the hTERT 5−12 containing the DNA construct on the surface of antidigoxigenin antibody−coated polystyrene beads, 0.1 ng (3.5×10^{-17} mol) of DNA was mixed with 1 μL of beads (2.10 μm in diameter, 0.5% w/v) in 10 μL of a 10 mM Tris buffer with 100 mM KCl (pH 7.4). The mixture was diluted to 1 mL with the same buffer after gently shaking at room temperature for 30 min. The DNA-immobilized beads were injected into a custom-made chamber and made ready for laser tweezers experiments.

We used home-built dual-trap 1064-nm laser tweezers to carry out the force−ramp assay[18,19] at 23 °C. One laser focus (mobile trap) grabbed the bead labeled with the DNA, while another focus (static trap) trapped a streptavidin-coated bead (1.87 μm diameter, Spherotech). The mobile trap, controlled by a movable mirror, brought the two beads close to each other, which allowed the tethering of the DNA construct between the two beads through affinity interactions. After this, the mobile trap moved two beads apart to a force of 60 pN, followed by relaxing to 0 pN at a load rate of 5.5 pN/s and a data collection rate of 1000 Hz.

For assays in the presence of a 15-nt hairpin complement, the hTERT 5−12 DNA construct was incubated with antidigoxigenin beads in the same way as mentioned above, except that 10 μM 15-nt hairpin complement was added into the incubation buffer. This concentration was maintained in the flow buffer during force−ramp assays in laser tweezers.

Plot of Change in Contour Length versus Force. Each extension−force plot was separated into a stretching (black) and a relaxing (red) region (Figure 1c). The extension during the stretching was subtracted from that during the relaxing at a particular force. The resulting change in extension (Δx) was then converted to the change in contour length (ΔL) through the worm-like-chain (WLC) model:[19,38,39]

$$\frac{x}{L} = 1 - \frac{1}{2}\left(\frac{k_B T}{FP}\right)^{1/2} + \frac{F}{S}$$

where x is the end-to-end distance (or extension), L is the contour length, F is the force, S is the stretching modulus (1226 pN), P is the persistent length (51.95 nm), k_B is the Boltzmann constant, and T is absolute temperature. In our case, the fully unfolded structure in the hTERT 5−12 is an ssDNA, which can be fit by a freely-jointed-chain (FJC) model.[40,41] However, our construct contains two dsDNA handles (4700 bp) with combined contour length 2 orders of magnitude longer than that of the 54-nt fragment (including the hTERT 5−12 sequence with flanking deoxythymidines); therefore, the dsDNA element should predominate in the equation that describes the entire DNA construct. A numerical calculation on a fitting equation based on the sequential WLC and FJC models indeed showed that the FJC component in our DNA construct only contributed less than 1% from 0 to 60 pN, which was the force range used in our experiments. Thus, the WLC model for dsDNA can be used to obtain an apparent contour length for the entire DNA construct. The difference in the apparent contour length between the DNA construct that contains a folded structure (the stretching curve before the rupture event) and that which contains an unfolded structure (the relaxing curve before the refolding event) represents the ΔL of the folded structure. A similar rationale has been used to calculate the ΔL of proteins in the single-molecule AFM experiments.[38]

To validate this new method, we performed the ΔL measurement using conventional procedures,[38] in which a single ΔL value was obtained from one set of F−X curves. As shown in Figure S1a (inset), the conventional method and the hysteresis region−based new analysis yielded identical ΔL values.

Kernel Density Estimation and Bootstrapping Analysis. The probability density for each transition between a folded species and the unfolded ssDNA, p, is estimated according to the following Gaussian kernel expression:[23]

$$p = \frac{1}{\sigma 2\pi}\exp\left(-\frac{(x - \Delta L)^2}{2\sigma^2}\right)$$

where ΔL and σ are the change in contour length for the transition and its associated standard error, respectively, and x is defined by the range $\Delta L \pm 3\sigma$. The σ was the average standard error for the two regions flanking the transition event (see Figure S10 for the histograms). The probability density for each step was then added to construct the kernel density estimation (Figure 2d). Bootstrapping analysis was performed by a random resampling procedure. The three highest peaks from each kernel density estimation were identified by an Igor6 program. After 3000 times of resampling, a histogram was constructed for these selected peaks (Figure 2e). This approach yielded the unfolding ΔL centered at 16.2 ± 0.2, 13.0 ± 0.1, 9.0 ± 0.2, and 4.1 ± 0.1 nm. For refolding, the ΔL values are centered at 16.0 ± 0.2, 13.0 ± 0.1, 9.0 ± 0.2, and 4.1 ± 0.2 nm, respectively (Figure 2e).

Calculation of the Nucleotides in a Folded Structure. On the basis of the transition between fully folded and fully unfolded conformations (ΔL = 16.1 ± 0.2 nm for the average of unfolding and refolding transitions, Figure 2e), the single nucleotide contour length, L_{sn}, was calculated to be 0.40 nm using the following:[5,19,37,42]

$$L_{sn} = \frac{\Delta L + x}{n}$$

where ΔL is the change in contour length, n is the number of nucleotides, and the end-to-end distance (x) is estimated to be 1 nm (PDB: 2KZD).[43] This L_{sn} is within the range of values obtained from our previous results and references (0.36−0.45 nm).[19,44,45] The same equation was used to calculate the number of nucleotides involved in an intermediate structure. For the unfolding transition with ΔL = 13.0 ± 0.1 nm (the corresponding refolding ΔL = 13.0 ± 0.1 nm), this equation reveals that 32 nucleotides are involved in the structure (n = 32) when L_{sn} = 0.40 nm and x = 1.2 nm (the distance between Hoogsteen G−G pairs[43]) are used in the calculation. The use of x = 1.2 nm is validated as the "32" intermediate (Figure 2f, right) has its two terminal guanine residues joined by the Hoogsteen pair. For the unfolding transitions with ΔL = 9.0 ± 0.2 and 4.1 ± 0.1 nm (the corresponding refolding ΔL values are 9.0 ± 0.2 and 4.1 ± 0.2 nm, respectively), calculations reveal 25 and 15 nucleotides in the structures, respectively, based on L_{sn} = 0.40 nm and x = 1.8 nm (the diameter of a dsDNA with Watson−Crick base pairs[46]). The use of x = 1.8 nm is again validated because both 25-nt and 15-nt intermediates have their ends joined by Watson−Crick base pairs (Figure 2f).

Calculation of Unfolding ($k_{f\to u}$) or Refolding ($k_{u\to f}$) Rate Constants. The calculation of the unfolding ($k_{f\to u}$) or refolding ($k_{u\to f}$) rate constant was based on the Evans model[47] (Figures S4 and S8):

$$\ln\left[r\ln\left[\frac{1}{N(F,r)}\right]\right] = \ln\left[\frac{k_{f\to u}}{(x^{\ddagger}_{f\to u}/k_B T)}\right] + \left(\frac{x^{\ddagger}_{f\to u}}{k_B T}\right)F$$

$$\ln\left[-r\ln\left[\frac{1}{U(F,r)}\right]\right] = \ln\left[\frac{k_{u\to f}}{(x^{\ddagger}_{u\to f}/k_B T)}\right] - \left(\frac{x^{\ddagger}_{u\to f}}{k_B T}\right)F$$

where r is the load rate (5.5 pN/s) and $N(F, r)$ and $U(F, r)$ represent the fraction of folded and unfolded populations for unfolding and refolding processes, respectively, at the specific force F and the load rate, $x^{\ddagger}$ is the distance from the folded state to the transition state, k_B is the Boltzmann constant, and T is absolute temperature.

3D Free Energy Landscape Construction. The change in free energy, ΔG, for a particular transition was calculated by the Crooks theorem (Figure S5):[27,48]

$$\frac{P_U(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G}{k_B T}\right)$$

where k_B is the Boltzmann constant, T is absolute temperature, and $P_U(W)$ and $P_R(-W)$ represent the probability of population with a work W for unfolding and refolding transitions, respectively. The work was calculated with the following equation:

$$W = \sum_{i=1}^{N_s} F_i \Delta x_i$$

where F and Δx represent the rupture force and the rupture distance, respectively, for a folded species.

The rupture force histogram of the transition from fully folded state (or "42") to fully unfolded state (or "0") was fitted to the Dudko model:[24]

$$p(F) \propto \frac{k(F)}{r} \exp\left[\frac{k_{off}}{x^{\ddagger}_r} - \frac{k(F)}{x^{\ddagger}_r}\left(1 - \frac{x^{\ddagger}F}{\Delta G^{\ddagger}}\nu\right)^{1-1/\nu}\right]$$

where $k(F) = k_{off}(1 - (x^{\ddagger}F/\Delta G^{\ddagger})\nu)^{(1/\nu)-1} \exp\{\Delta G^{\ddagger}[1 - (1 - (x^{\ddagger}F/\Delta G^{\ddagger})\nu)^{1/\nu}]\}$, k_{off} is the unfolding rate constant at zero force, $x^{\ddagger}$ is the activation distance from the folded state to the transition state, $\Delta G^{\ddagger}$ is the activation energy, and ν is the factor depicting the shape of the energy barrier ($\nu = 1/2$ for a sharp, cusp-like shape barrier, while $\nu = 2/3$ for a softer, cubic shape). The average from the $\nu = 1/2$ and $\nu = 2/3$ fittings was used to construct the energy landscape in Figure 3d, Table 1, and Figure S6.

The activation energy, $\Delta G^{\ddagger}$, for other major transitions was calculated by the Arrhenius equation:

$$k = A \exp\left(\frac{-\Delta G^{\ddagger}}{k_B T}\right)$$

where k is the rate constant obtained either by the Dudko model or by the Evans model discussed above, k_B is the Boltzmann constant, and T is absolute temperature. The prefactor, $A = 5.03 \times 10^{19}$ s^{-1}, was obtained from the $\Delta G^{\ddagger}$ of the "42"→"0" unfolding process by the Dudko model discussed above. For unfolding of the intermediates that resemble hairpins, the prefactor[49] of $A = 10^5$ s^{-1} was used to obtain the activation energy (Figure 3d, Table 1, and Figure S6) .

Unfolding angles were calculated according to the proposed structure in the hTERT 5−12 fragment[20] (Figure 1a). NMR structures (PDB 2KZD and PDB 1NGU)[43,46] were used to represent the G-quadruplex and the hairpin elements in the hTERT, respectively. First, coordinates of the two phosphate atoms flanking the two ends of a particular species (either 42, 32, 25, or 15 nt) were projected onto a $y-z$ plane. For example, coordinates for the fully folded species (42-nt) were obtained from P31 and P617 atoms in PDB 2KZD. After projection onto the $y-z$ plane, coordinates of (5.016, −9.293) and (−3.040, −14.274) were obtained. The line between these two points was then designated as the 0° angle and served as a reference. After the two ends of other species were projected onto the same plane, the unfolding angle of the species was measured by the angle between its projected line and the reference line. To orientate the hairpin (PDB 1NGU)[46] and the G-quadruplex (PDB 2KZD)[43] in the structure described in Figure 1a, we adjusted the angle of the 25-nt species (measured from the hairpin structure) to the same value of the 25-nt intermediate measured from the G-quadruplex structure. This adjustment was applied to determine the unfolding angle of the 15-nt species measured from the hairpin (PDB 1NGU).[46]

The angular, kinetic, and energetic values are summarized in Table 1.

DMS Footprinting. PAGE-purified oligonucleotides were 5′-end ^{32}P labeled, diluted to 1 μM, and annealed in the presence or absence of varying concentrations of 15-nts by heating at 70 °C and slow cooling over 2 h. A 10% native polyacrylamide gel was used to separate the annealed and linear species, which were excised and extracted using 10 mM Tris-HCl. In the presence of 100 mM KCl, the G-quadruplex was induced by heating to 65 °C for 10 min and slow cooling over 2 h. Two micrograms of calf thymus DNA was added to each sample, and the samples were treated with 0.5% DMS as a final concentration for 10 min at room temperature. Each reaction was quenched by adding 3.5 μL of β-mercaptoethanol, and 10 μL of a 10% glycerol solution was added prior to running on a 10% native polyacrylamide gel. The unimolecular oligonucleotide species were separated by EMSA, excised, extracted in dH_2O, and precipitated at −20 °C in the presence of 74% ethanol, 100 mM NaOAc, and 6.5 μg/mL yeast tRNA as final concentrations. The oligonucleotides were then subjected to treatment with 20% piperidine (v/v in 10 mM Tris-HCl) at 90 °C for 18 min, and the resulting cleavage products were separated on a 16% sequencing gel. The protection of each residue against DMS treatment for samples in the buffer with 100 mM KCl was evaluated by comparing the gel density of a band with that of the corresponding band for samples in the buffer without KCl.[20]

Molecular Modeling. Structures of the 25-nt species (Figure 2f) with and without the 15-nt complementary sequence (Figure 4b) were constructed using the Insight II/Biopolymer modeling program.[50] Charges were assigned using the consistent valence force field (CVFF). Sodium counterions were added to the structures, and the entire system was soaked in a 10 Å layer of TIP3P water molecules. The structures were then minimized separately using 30 000 steps of Discover 3.0 minimization within Insight II. This was followed by molecular dynamics simulations with 50 ps equilibration and 950 ps simulations at 300 K. Frames were collected every 1000 fs during the simulation. Trajectories were analyzed on the basis of potential energy. Average structure was created using 20 lowest potential energy frames, and this average structure was refined using 100 000 steps of Discover 3.0 minimization. These minimized average structures were then used for modeling studies. The complex between hTERT 25-nt and its complementary 15-nt sequence was created using interactive docking and was refined using a similar procedure as described above.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S10. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

hurley@pharmacy.arizona.edu; hmao@kent.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

H.M. is thankful for support from the NIH (DK081191-01) and NSF (CHE-1026532). L.H.H. acknowledges support from the NIH (GM085585) and the National Foundation for Cancer Research (VONHOFF0601). We thank Dr. Tracy Brooks for helpful discussions in processing the data. We are grateful to Dr. David Bishop for proofreading and editing the text for submission.

■ REFERENCES

(1) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. *Annu. Rev. Biophys.* **2010**, *39*, 159−83.

(2) Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S. *Science* **2005**, *309*, 2057−2060.

(3) Peng, Q.; Li, H. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1885−1890.

(4) Onoa, B.; Dumont, S.; Liphardt, J.; Smith, S. B.; Tinoco, I. Jr.; Bustamante, C. *Science* **2003**, *299*, 1892−1895.

(5) Greenleaf, W. J.; Frieda, K. L.; Foster, D. A. N.; Woodside, M. T.; Block, S. M. *Science* **2008**, *319*, 630−633.

(6) Li, P. T.; Vieregg, J.; Tinoco, I. Jr. *Annu. Rev. Biochem.* **2008**, *77*, 77−100.

(7) Brooks, T. A.; Kendrick, S.; Hurley, L. H. *FEBS J.* **2010**, *277*, 3459−3469.
(8) Brooks, T. A.; Hurley, L. H. *Nat. Rev. Cancer* **2009**, *9*, 849−861.
(9) Paeschke, K.; Capra, J. A.; Zakian, V. A. *Cell* **2011**, *145*, 678−691.
(10) Decorsière, A.; Cayrel, A.; S., V.; Millevoi, S. *Genes Dev.* **2011**, *25*, 220−225.
(11) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 261−275.
(12) Qin, Y.; Hurley, L. H. *Biochimie* **2008**, *90*, 1149−1171.
(13) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908−2916.
(14) Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2006**, *128*, 15461−15468.
(15) Petraccone, L.; Trent, J. O.; Chaires, J. B. *J. Am. Chem. Soc.* **2008**, *130*, 16530−16532.
(16) Schonhoft, J. D.; Bajracharya, R.; Dhakal, S.; Yu, Z.; Mao, H.; Basu, S. *Nucleic Acids Res.* **2009**, *37*, 3310−3320.
(17) Thirumalai, D.; Woodson, S. A. *Acc. Chem. Res.* **1996**, *29*, 433−439.
(18) Mao, H.; Luchette, P. *Sens. Actuators, B* **2008**, *129*, 764−771.
(19) Yu, Z.; Schonhoft, J. D.; Dhakal, S.; Bajracharya, R.; Hegde, R.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2009**, *131*, 1876−1882.
(20) Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 10878−10891.
(21) Shakhnovich, E. *Chem. Rev.* **2006**, *106*, 1559−1588.
(22) Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Herschlag, D.; Block, S. M. *Science* **2006**, *314*, 1001−1004.
(23) Cheng, W.; Arunajadai, S. G.; Moffitt, J. R.; Tinoco, I. Jr.; Bustamante, C. *Science* **2011**, *333*, 1746−9.
(24) Dudko, O.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96*, 108101.
(25) Evans, E. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 105−28.
(26) Risitano, A.; Fox, K. R. *Nucleic Acids Res.* **2004**, *32*, 2598−2606.
(27) Crooks, G. E. *Phys. Rev. E* **1999**, *60*, 2721−2726.
(28) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I. J.; Bustamante, C. *Nature* **2005**, *437*, 231−234.
(29) Shakhnovich, E. *Chem. Rev.* **2006**, *106*, 1559−88.
(30) Shakhnovich, E. I. *Phys. Rev. Lett.* **1994**, *72*, 3907−3910.
(31) Maddox, J. *Nature* **1994**, *370*, 13.
(32) Go, N.; Taketomi, H. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 559−63.
(33) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *J. Mol. Biol.* **1995**, *252*, 460−71.
(34) Govindarajan, S.; Goldstein, R. A. *Proteins* **1995**, *22*, 413−8.
(35) Zuo, G.; Wang, J.; Wang, W. *Proteins* **2006**, *63*, 165−73.
(36) Urquidi, V.; Tarin, D.; Goodison, S. *Annu. Rev. Med.* **2000**, *51*, 65−79.
(37) Dhakal, S.; Schonhoft, J. D.; Koirala, D.; Yu, Z.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2010**, *132*, 8991−8997.
(38) Carrion-Vazquez, M.; Oberhauser, A. F.; Fisher, T. E.; Marszalek, P.; Li, H.; Fernandez, J. M. *Prog. Biophys. Mol. Biol.* **2000**, *74*, 63−91.
(39) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185−6190.
(40) Smith, S. B.; Cui, Y. J.; Bustamante, C. *Science* **1996**, *271*, 795−799.
(41) Dessinges, M. N.; Maier, B.; Zhang, Y.; Peliti, M.; Bensimon, D.; Croquette, V. *Phys. Rev. Lett.* **2002**, *89*, 248102.
(42) Dietz, H.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16192−16197.
(43) Lim, K. W.; Lacroix, L.; Yue, D. J.; Lim, J. K.; Lim, J. M.; Phan, A. T. *J. Am. Chem. Soc.* **2010**, *132*, 12331−42.
(44) Mills, J. B.; Vacano, E.; Hagerman, P. J. *J. Mol. Biol.* **1999**, *285*, 245−257.
(45) Laurence, T. A.; Kong, X.; Jager, M.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17348−17353.
(46) Shiflett, P. R.; Taylor-McCabe, K. J.; Michalczyk, R.; Silks, L. A.; Gupta, G. *Biochemistry* **2003**, *42*, 6078−89.
(47) Li, P. T.; Collin, D.; Smith, S. B.; Bustamante, C.; Tinoco, I. Jr. *Biophys. J.* **2006**, *90*, 250−60.
(48) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I. Jr.; Bustamante, C. *Nature* **2005**, *437*, 231−4.
(49) Woodside, M. T.; Behnke-Parks, W. M.; Larizadeh, K.; Travers, K.; Herschlag, D.; Block, S. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6190−6195.
(50) *Insight II 2005L, Molecular Modeling Software*; Accelrys Inc.: San Diego, CA.