JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

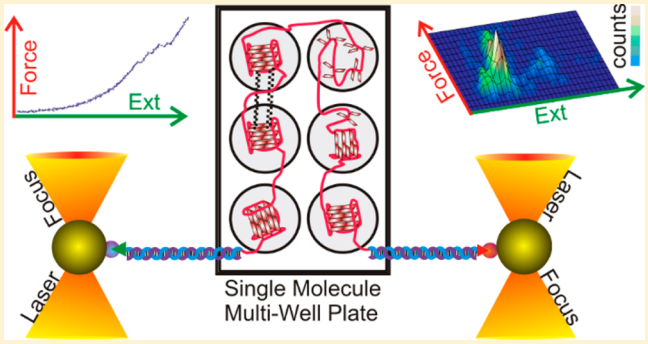
Interaction of G-Quadruplexes in the Full-Length 3′ Human Telomeric Overhang

Jibin Abraham Punnoose, Yunxi Cui, Deepak Koirala, Philip M. Yangyuoru, Chiran Ghimire, Prakash Shrestha, and Hanbin Mao*

Department of Chemistry and Biochemistry, Kent State University, Kent, Ohio 44242, United States

Supporting Information

ABSTRACT: The 3′ human telomeric overhang provides ample opportunities for the formation and interaction of G-quadruplexes, which have shown impacts on many biological functions including telomerase activities in the telomere region. However, in the few investigations on DNA constructs that approach to the full length of the human telomeric overhang, the presence of higher-order quadruplex−quadruplex interactions is still a subject of debate. Herein, we employed dynamic splint ligation (DSL) to prepare a DNA construct, 5′-$(TTAGGG)_{24}$ or 24G, which has the length comparable to the full stretch of 3′ human telomeric overhang. Using mechanical unfolding assays in laser tweezers, we observed a minor population (∼5%) of higher-order interactions between G-quadruplexes, while the majority of the quadruplexes follow the bead-on-a-string model. Analyses on the noninteracting G-quadruplexes in the 24G construct showed features similar to those of the stand-alone G-quadruplexes in the 5′-$(TTAGGG)_4$ (4G) construct. As each 24G construct contains as many as six G-quadruplexes, this method offers increased throughput for the time-consuming mechanical unfolding experiments of non-B DNA structures.



INTRODUCTION

Eukaryotic chromosomes are terminated with telomeres which contain tandem repeats of Guanine (G)-rich nucleotide sequences. Telomere guards the chromosome from double strand breakage, chromosome fusion, and nuclease activities.[1] In human somatic cells, telomere is approximately 8−12 kb long, which becomes shorter on every cell division due to the end replication problem.[2,3] Shortening length of the telomere is associated with cell aging and death. In germ cells, most of the stem cells, and cancer cells, the enzyme telomerase reverse transcriptase is expressed to recognize and elongate the single-stranded overhang at the 3′ end of the telomere.[4] In human cells, this 3′ overhang has a length of 200 ± 75 nucleotides with 20−45 repeats of a consensus sequence, 5′-TTAGGG.[5] In the presence of monovalent metal ions such as Na^+ or K^+, four of these G-rich repeats can form a G-quadruplex structure by stacking three G-quartet layers, each of which contains four coplanar guanines stabilized by Hoogsteen hydrogen bonds.[6] It has been found that the G-quadruplex can function as a regulatory switch for the telomere elongation by interfering with telomerase activities.[7−9] Thus, structural information on the G-quadruplexes and their possible higher order interactions in the 3′ telomeric overhang is vital for designing highly specific cancer therapeutic agents.[10−12]

Most of current studies are focused on telomeric DNA with four TTAGGG repeats that can only host one G-quadruplex unit. The simplicity of the system has enabled higher resolution structural information on telomeric G-quadruplexes by NMR or X-ray techniques.[13,14] However, such a system oversimplifies the physiological context in which a telomere overhang may contain as many as 11 copies of G-quadruplexes. To the best of our knowledge, none of the studies has used a template in the range of full-length human telomeric overhang to investigate the G-quadruplex interactions. Only few investigations have focused on the neighboring effect of G-quadruplexes[15−17] which may present different biological effects than individual G-quadruplexes. When the length of G-rich DNA is long, the solid phase synthesis of deoxyoligonucleotides becomes challenging. In addition, there exists a high level of complexity in data analysis due to the structural polymorphism exhibited by telomeric G-quadruplexes or their intermediates, such as G-hairpin, G-Triplex and misfolded long loop G-quadruplex formed during the folding or unfolding of a G-quadruplex.[18−20] Due to these complexities, ensemble-average based methods present a degree of uncertainty to characterize quadruplexes in long telomeric sequences. For example, it remains controversial whether or not higher-order interactions between neighboring G-quadruplexes exist. While Sugimoto group reported a "bead-on-a-string" model to describe the absence of higher-order G-quadruplex interactions, others found the evidence of quadruplex-quadruplex interactions in telomeric DNA and

Received: September 30, 2014
Published: December 1, 2014

   dx.doi.org/10.1021/ja510079u | J. Am. Chem. Soc. 2014, 136, 18062−18069

RNA fragments using techniques such as X-ray crystallography, NMR, and mass spectrometry.[16,17,21−28]

Compared to ensemble assays, single-molecule methods are appropriate to investigate interactions among multiple G-quadruplexes since these methods are capable of probing individual species one at a time in a reaction mixture. In addition, the nanomolar template concentration employed in single-molecule approaches allows one to investigate the property of G-quadruplex under conditions physiologically more relevant compared to ensemble average methods, in which micro- to millimolar concentrations are used.[29] Here, we have used mechanical unfolding in optical tweezers instrument[30] to investigate the interactions among a maximum of six telomeric G-quadruplexes in a DNA sequence of 24 TTAGGG repeats (24G). We have innovated a dynamic splint ligation (DSL) strategy to prepare this long G-rich DNA, whose length falls into the range for a full-length 3′ human telomeric overhang. Our results have revealed that while G-quadruplexes do not interact with each other in the DNA sequences that can form less than six G-quadruplexes, there is a minor population of higher-order interactions (5%) in the 24G DNA construct that can host six G-quadruplexes. As the mechanical unfolding force, the size, and the free energy of unfolding are indistinguishable for noninteracting G-quadruplexes formed in DNA constructs with variable lengths, the 24G construct allows one to probe DNA secondary structures with a much increased throughput at the single molecular level.

MATERIALS AND METHODS

Materials. DNA oligomers were purchased from Integrated DNA Technologies (www.idtdna.com) and further purified by denaturing PAGE gel and stored at −20 °C. Enzymes required for the preparation of DNA constructs were purchased from NEB (www.neb.com). The polystyrene beads coated with streptavidin or antidigoxigenin for the single-molecule experiments were purchased from Spherotech (Lake Forest, IL).

Preparation of DNA Construct. DNA constructs containing 4, 8, and 12 G-tracts (see Supporting Information Table S1 for sequences) for mechanical unfolding/refolding experiments were synthesized using protocols described elsewhere.[15,31] Briefly, a DNA construct containing a single-stranded DNA oligomer with a sequence of 5′-$(TTAGGG)_4$, 5′-$(TTAGGG)_8$, and 5′-$(TTAGGG)_{12}$ was sandwiched between a biotin-labeled 2028 bp and a digoxigenin labeled 2690 bp dsDNA handles. These terminal labeling allows tethering of the DNA molecule between the streptavidin and anti-digoxigenin coated polystyrene beads, respectively.

Dynamic Splint Ligation (DSL). The oligomer containing 24G tracts was not commercially available. To prepare this construct, we started with commercially available oligonucleotides 12G1 and 12G2 (see Supporting Information Table S1 for sequences). These two sequences were ligated to the biotin-labeled 2028-bp fragment (prepared from the 629-2691th site in the pBR322 plasmid by PCR using a biotinlyated primer and a primer with an XbaI recognition site) and the digoxigenin-labeled 2690-bp fragment (prepared from the pEGFP plasmid through sequential digestions using SacI and EagI restriction enzymes), respectively. A splint ligation strategy was then applied to ligate these two long DNA pieces. To avoid the self-ligation of either 12G1 or 12G2 DNA, the 5′ end of the 12G1 and the 3′ end of the 12G2 were designed to mismatch with the splint DNA, which is a 24-nucleotide fragment (5′-TAACCCTAACCCTAACCCTA-ACCC) that can piece together 12G1 and 12G2 sequences through base pairing. To increase the ligation efficiency, the ligation conditions such as temperature, time, and ratio of the split DNA to the ligation substrates were optimized using the 12G1 and 12G2 fragments without their respective attachments to the costly biotin-labeled 2028 bp and digoxigenin-labeled 2690 bp handles. Optimization was performed by varying the ratio of splint DNA to template DNA from 1:1 to 6:1 (Figure 3B). Ligation time was varied from 6 to 48 h. Temperature ramping cycles (22 °C for 20 min and then 45 °C for 2 min) were performed to recycle the splints to increase the ligation efficiency. The yields of these ligation conditions were compared to those of the ligation performed under constant temperatures.

From the optimization, the best condition gave a ligation yield of 77% in 24 h at 1:6 template:splint ratio. Extending the time to 48 h did not improve the efficiency. Without temperature cycling, the yield was 63%. Although modest, the temperature cycling indeed helps to increase the ligation efficiency by recycling the splint oligos.

These optimized conditions were used for the synthesis of the DNA construct sandwiched between the biotin-labeled 2028 bp and the digoxigenin-labeled 2690 bp dsDNA. Before the mechanical unfolding experiments, the splint DNA was removed by annealing the ligated construct with 20 times excess of the oligomer complementary to the splint (Splr1 in Supporting Information Table S1). The removal of the splint was confirmed by the native gel shift assay (Supporting Information Figure S1).

Laser Tweezers Instrumentation. The single molecule investigation was carried out in a home-built laser tweezers instrument for which the detailed description has been reported previously.[30,32] All the experiments, unless specified otherwise, were carried out in a 10 mM Tris/100 mM KCl buffer, pH 7.4 at 23 °C. To start the single molecule experiments, digoxigenin-labeled DNA construct was immobilized onto a 2.10 μm polystyrene bead coated with anti-digoxigenin via digoxigenin−anti-digoxigenin antibody interaction. The DNA immobilized bead and the streptavidin coated bead were trapped by two laser foci and the DNA construct was then tethered between these two beads. In a typical force−extension experiment, the tethered DNA was extended below the plateau force (maximum 60 pN) and relaxed to 0 pN by moving one of the trapped beads with a loading rate of 5.5 pN/s. The force−extension (F−X) curves were recorded at 1 kHz using a LabView program (National Instruments Corp., Austin, TX). These raw data were filtered with a Savitzky−Golay function with a time constant of 10 ms using a Matlab program (The MathWorks, Natick, MA).

Data Analysis. The change in extension (Δx) at a particular force (F) was calculated as the extension difference between the stretching and the relaxing traces at that force. The resulting Δx at this force was then converted to the change in contour length (ΔL) using the following wormlike-chain (WLC) model,[33,34]

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

where Δx is the change in end-to-end distance (or extension) between the two optically trapped beads, ΔL is the change in contour length, k_B is the Boltzmann constant, T is the absolute temperature, P is the persistent length of dsDNA (50.8 nm),[29] and S is the stretching modulus (1243 pN).[29]

Prediction of the Probability of Simultaneous Unfolding of Two G-Quadruplexes. Given the probability of unfolding a single structure at a particular force F_i as $P(F_i)$, then the probability of unfolding two structures simultaneously ($P_{simultaneous}$) between the observed minimum (F_1) and maximum (F_2) unfolding forces can be calculated by eq 2,[35]

$$P_{simultaneous} = \sum_{i=F_1}^{F_2} P(F_i)^2 \quad (2)$$

Percent Formation of Secondary Structures. Probabilities of secondary structures formed in different constructs (P) were calculated according to the formula,

$$P = \frac{N}{N_{total}} \times 100\% \quad (3)$$

Where N is the number of unfolding features observed in a total of N_{total} F−X curves.

The percentage of G-quadruplex formed among secondary structures in a particular construct was determined through multipeak

dx.doi.org/10.1021/ja510079u | J. Am. Chem. Soc. 2014, 136, 18062−18069

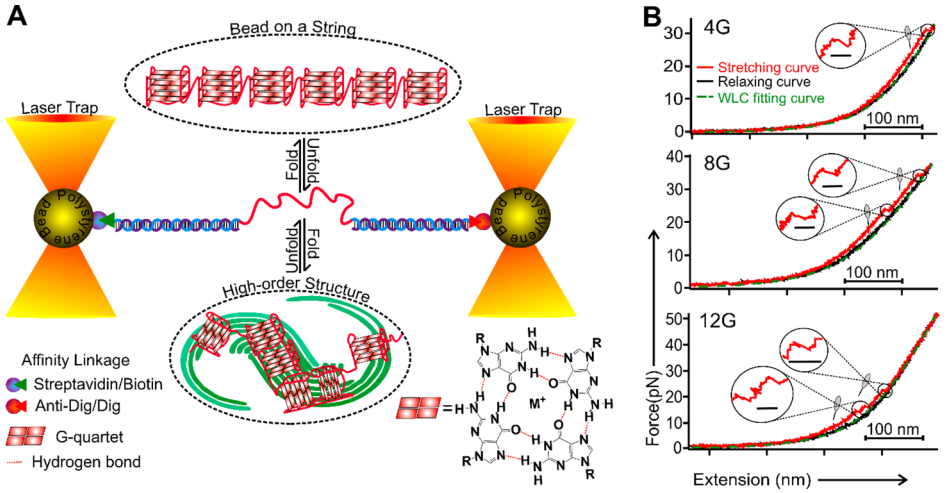


Figure 1. (A) Experimental setup for mechanical unfolding experiments. The single-stranded telomeric DNA is sandwiched between a biotin-labeled 2028 bp DNA and a digoxigenin-labeled 2690 bp DNA, which are tethered to streptavidin and anti-digoxigenin coated polystyrene beads, respectively. (B) Sequential unfolding of 1, 2, and 3 G-quadruplexes in the 4G, 8G, and 12G DNA constructs, respectively. Scale bars represent 10 nm.

Gaussian fitting in Figures 2A and 4A using Igor Pro (WaveMetrics, Portland, OR) software.

Hierarchical Cluster Analysis (HCA). To differentiate the higher-order G-quadruplex structures from the noninteracting G-quadruplexes in the 24G DNA construct, we introduced the hierarchical cluster assay (HCA).[36] Different from other discriminant assay such as Linear Discriminant Analysis (LDA, see below), the HCA does not require training to categorize groups in a sample.[37] Therefore, it is ideal to classify samples into various groups without prior knowledge of their identities.

In our HCA, unfolding features were assigned into clusters in a dendrogram (Supporting Information Figure S4) according to the relative distance calculated using different variables such as rupture force or unfolding work of each individual sample. For a valid HCA, the size of the sample (n) and the number of variables (m) should follow the Formann formula,[38] $n > 2^m$. In our analysis, we applied 3 variables ($m = 3$): the rupture force (F), the change-in-contour length (ΔL), and the work of the unfolding.[15] Since the number of total samples (unfolding features) is $n = 1118$, it is much larger than the $2^m = 2^3 = 8$. We used Euclidean distance (d)[39] to define the difference between sample i and sample j according to eq 4,

$$d = \sqrt{(F_i - F_j)^2 + (L_i - L_j)^2 + (W_i - W_j)^2} \quad (4)$$

Matlab programs[40] were used to perform HCA on the 24G construct in a two-step process. In the first step, the distance calculated from eq 4 was used to categorize all the data points into two groups (Supporting Information Figure S4A). While the Group 2 consists of higher-order structures (Figure 5B, green ellipse) as evidenced by high rupture force and high ΔL values (see main text), the Group 1 is a combination of partially folded structure, noninteracting G-quadruplexes, and simultaneous unfolding of two G-quadruplex by chance. To distinguish these populations, we performed a second-step HCA to categorize the data in Group 1 into 3 subgroups using ΔL as the distance (Supporting Information Figure S4B). The resultant 4 populations in the 24G construct were plotted in Figure 5B, in which partially folded structures (brown), G-quadruplexes (red), simultaneous unfolding of two noninteracting G-quadruplexes (blue), and higher-order G-quadruplexes (green) are classified at the 90% confidence level (see Figure 5B).

Linear Discriminant Analysis (LDA). LDA is a statistical method for the assignment of unclassified samples to appropriate classes determined previously. Briefly, a discriminant function is first calculated based on the covariance matrices of the variables of the training objects (the known sets).[37] This function can maximize the distance obtained from the variables between different known classes. For an unclassified sample, the same discriminant function is employed to estimate the distances between the unclassified sample and the data points in the known classes. The assignment of the unclassified sample is determined by the shortest distance between the sample and the designated class.[37] We used Matlab LDA program and employed ΔL and rupture force F as discriminant variables to classify the populations in the 4G–12G constructs (Figure 5B) according to the four populations determined in the 24G construct by the HCA (see above).

Calculation of the Change in Free Energy of Unfolding (ΔG_{unfold}). The stretching and relaxing force–extension curves were used to calculate the change in free energy (ΔG_{unfold}) associated with the unfolding of the G-quadruplex according to Jarzynski's equation[41] for nonequilibrium systems,

$$\Delta G_{unfold} = -k_B T \ln \sum_{i=1}^{N} \frac{1}{N} e^{-(W_i/k_B T)} \quad (5)$$

Where N is the number of unfolding events and W_i is the nonequilibrium work done for unfolding each structure.

Prior to the determination of the ΔG_{unfold} for noninteracting G-quadruplexes using eq 5, G-quadruplex population was deconvoluted from the partially folded structure (G-Triplex) or higher-order structures. First, the percentage population for each species in a particular construct was determined using multipeak Gaussian fitting of ΔL histograms (Figures 2A and 4A). The overlapping populations between any two species were then randomly assigned to each of the species according to the ratio between the two species in each bin of the particular ΔL histogram.[31] To calculate ΔG_{unfold} for higher-order G-quadruplexes, we applied eq 5 for the green ellipse determined by the HCA (see above) in Figure 5B.

RESULTS AND DISCUSSION

G-Quadruplex Units in the 8G or 12G Telomeric Fragment Follow the Bead-on-a-String Model. Using optical tweezers, we performed force-ramp assays (see Materials and Methods and Figure 1) to mechanically unfold and refold the DNA secondary structures formed in human telomeric fragments, 5′-$(TTAGGG)_n$, where n = 4 (4G), 8 (8G), and 12 (12G) indicate the maximal formation of 1, 2, and 3 G-quadruplex units, respectively. These telomeric fragments were sandwiched between a biotin-labeled 2028 bp dsDNA and a digoxigenin-labeled 2690 bp dsDNA for their attachments to the two optically trapped polystyrene particles (Figure 1A). In a

dx.doi.org/10.1021/ja510079u | J. Am. Chem. Soc. 2014, 136, 18062–18069

typical experiment, force ramping of 5.5 pN/s was carried out to stretch the DNA construct until the unfolding of the structures was observed, which is indicated by a sudden change in the extension or force in a force−extension (F−X) curve (see Figure 1B for typical F−X curves of 4G, 8G, and 12G constructs, see Supporting Information Figure S2A for the rest). By the Worm-Like-Chain (WLC)[33] model that fits well to the F−X curves (Figure 1B), we can retrieve the size of folded structure as change-in-contour length (ΔL) and the mechanical stability of the structure by recording the rupture force ($F_{rupture}$) at which unfolding occurs.[42] To allow the complete unfolding of all structures in telomeric fragments, the F−X curves were recorded up to 60 pN before relaxing to 0 pN to refold DNA structures within a 30-s incubation period.[43]

For the telomeric fragment 4G, the ΔL histogram revealed two major populations of 9.0 ± 0.2 nm and 5.3 ± 0.3 nm in ΔL (Figure 2A). Given the end-to-end distances (x) of 1.1 ± 0.2 nm (for hybrid-1 telomeric G-quadruplex) and 1.5 ± 0.1 nm (for partially folded structure (G-triplex)),[44] these ΔL values give 22 ± 1 and 15 ± 1 nucleotides (nts) in the folded structures according to the relationship, nts = $(\Delta L + x)/L_{nt}$, where L_{nt} is the contour length per nucleotide (0.45 nm) determined from literature.[42,45] These values are consistent with the number of nucleotides involved in a fully folded telomeric G-quadruplex and a partly folded structure,[44,46] respectively. In addition to these two species, the ΔL histograms of 8G and 12G showed an additional ΔL population of 16.9 nm. This species can be simultaneous unfolding of two G-quadruplexes (Supporting Information Figure S2A) either by chance or due to the disruption of the higher-order interactions between the quadruplexes.[35] To determine the origin of this population, we estimated the probability of simultaneous unfolding of two unrelated G-quadruplexes. To this end, we first obtained the probability of unfolding G-quadruplexes in the 4G fragment from the rupture force histogram.

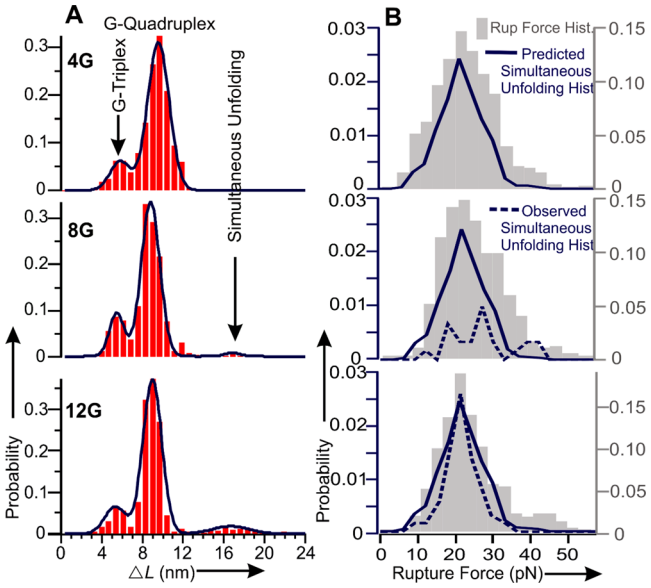


Figure 2. Mechanical unfolding of secondary structures formed in the 4G, 8G, and 12G DNA constructs. (A) Change-in-contour-length (ΔL) histograms. (B) Rupture force histograms (gray bars). Predicted histograms for simultaneous unfolding are shown in solid blue trace, while the observed simultaneous unfolding are shown in dotted blue trace.

Simultaneous unfolding of the two G-quadruplexes at a particular force was then determined by squared probabilities[35] (see Materials and Methods). Integration of probabilities in the force range of (0−60) pN revealed 10.4 ± 0.2% for simultaneous unfolding of two G-quadruplexes by coincidence. In comparison, the observed percentages of simultaneous unfolding are 3.5 ± 0.7% and 6.7 ± 0.5%, respectively, in the 8G and 12G fragments (Figure 2). These results suggest the absence of quadruplex-quadruplex interactions in these two telomeric fragments. Previously, the lack of higher-order interaction observed between G-quadruplexes has led Sugimoto and co-workers to propose a bead-on-a-string model for G-quadruplexes in the 3′ human telomeric overhang region.[16] Our results support this model for 8G and 12G telomeric fragments.

Although the observed probabilities of simultaneous unfolding were less than that predicted by statistical coincidence, we observed that the discrepancy becomes smaller when the length of the telomeric fragments increases (compare the middle and bottom panels in Figure 2B). If this trend continues, it suggests that higher-order G-quadruplex interactions may be present as telomeric fragments become longer. In human cells, the 200 ± 75 nt long telomeric overhang can host 5 to 11 G-quadruplexes. It is of high physiological significance to examine whether or not higher-order G-quadruplex interactions exist in longer telomeric sequences, which may bring different biological consequences with respect to the bead-on-a-string model. For example, it is anticipated that extra stability conveyed by the quadruplex-quadruplex interaction may alter the activities of telomerase and associated proteins during the telomeric elongation processes.

higher-order Structures Exist As a Minor Population in the 24G Telomeric Fragment. To investigate the possible quadruplex−quadruplex interaction in the longer telomeric fragment, we prepared a DNA construct, 5′-$(TTAGGG)_{24}$, which has the length comparable to the full-length telomeric overhang. Due to the presence of multiple runs of guanine repeats, the synthesis of this construct becomes challenging. We modified a splint ligation process in which two pieces of ssDNA (12G1 and 12G2, see Supporting Information Table S1) were first annealed in the presence of a splint oligo spl1, and then ligated using T4 DNA ligase (Figure 3A). Compared to regularly used splint strategy,[47] two innovations were used in our approach. First, to facilitate the removal of the splint after ligation, we used shorter splint sequences (Supporting Information Table S1). However, this introduced a new problem as majority of the splint DNA are consumed by hybridization with the repetitive complementary sequences in the two single-stranded 12G1 and 12G2 substrates. As a result, there are little splint fragments available to bring the two substrates together, which is thermodynamically unfavorable. To circumvent this issue, we introduced a second innovation, temperature cycling. By raising the temperature close to the melting of the splint-substrate duplex (45 °C), we disrupted the hybridization so that short splints become available to bring the two substrates together. We then reduced the temperature to settle this transient ternary complex, priming for the splint ligation by T4 DNA ligase at 22 °C. As the ligation reaction is almost irreversible, by multiple cycles of this temperature process, we were able to accomplish the ligation effectively. After ligation, the splint DNA was removed after addition of a complementary DNA (splr1) in excess (see Supporting Information Table S1 for sequences and Figure 3A,B and Supporting Information Figure S1 for synthesis). Since the

dx.doi.org/10.1021/ja510079u | J. Am. Chem. Soc. 2014, 136, 18062−18069

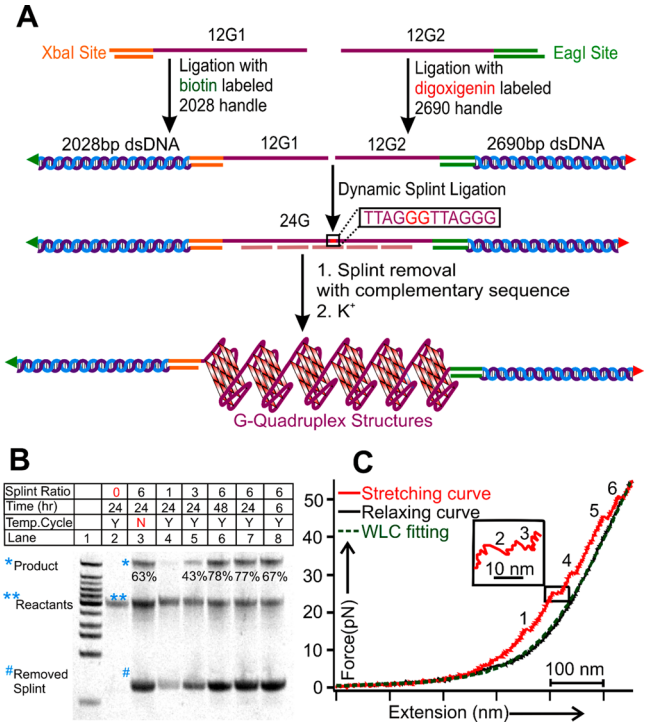


Figure 3. Synthesis strategy and characterization of the 24G construct. (A) Dynamic splint ligation of the 24G construct. (B) Native PAGE gel for the splint ligation of the 24G construct. Splint vs substrate ratio, time, and temperature cycling were varied for optimization. (C) Observed unfolding of six G-quadruplex units in the 24G construct. The green dotted curve represents WLC fitting.

splint DNA participates into the splint ligation with a dynamic on−off fashion, we named the process as the Dynamic Splint Ligation or DSL. After optimization of the ratio of the splint vs template DNA as 6:1, 78% ligation yield was obtained with complete removal of the splint DNA. We expect this DSL method is instrumental for the preparation of long fragments that contain multiple nucleic acid repeats, especially G-rich repeats that are difficult to synthesize. After synthesis, 24G fragment was tethered between the two optically trapped beads and force-ramping experiments were carried out as described above. As shown in Figure 3C, we observed up to six unfolding features, each of which has ∼9 nm ΔL, a value consistent with the size of a G-quadruplex observed above. When we mechanically unfolded the 24G construct in a buffer without metal ions or with 100 mM Li^+, few rupture events with the same ΔL range (2.3−11.0 nm) were observed (6.6 and 7.1%, respectively. See Supporting Information Figure S6). In comparison, 329% rupture events (average of 3.3 transitions per F-X curve) were observed in the buffer with 100 mM K^+ (see below). Other control experiments on constructs with 2 or 3 G-tracts also yielded features with reduced percentage of formation (7.1% and 20% for the 2G and 3G constructs, respectively, in a 10 mM Tris buffer (pH 7.4) with 100 mM K^+; 2% for the 3G construct in a 10 mM Tris buffer (pH 7.4), without ions, see Supporting Information Figure S6) . These observations validate the successful synthesis of the 24G fragment in which a maximum of 6 G-quadruplex structures can form.

Analyses of 351 F−X curves from 27 molecules of the 24G fragment allowed us to obtain histograms of ΔL measured during individual unfolding events (Figure 4A). Similar to those

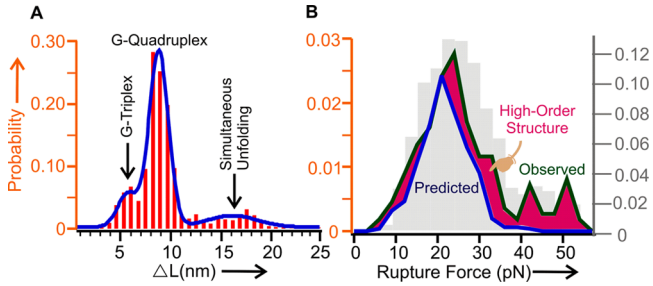


Figure 4. (A) ΔL histogram of structures in the 24G construct. (B) Comparison of predicted (blue profile trace) and observed (green profile trace) simultaneous unfolding events. The rupture force histogram of the G-quadruplex in the 4G fragment is plotted in gray background. See detailed rupture force histogram in Supporting Information Figure S3.

observed for the 8G and 12G fragments, we found the majority (73 ± 2%) of the structure is G-quadruplex with ΔL = 8.7 ± 0.1 nm, whereas partially folded structure corresponding to the length of G-Triplex (ΔL = 5.7 ± 0.1 nm) is present as a minor population (12.2 ± 1.8%). Interestingly, the probability of the population with large ΔL (16.4 ± 0.8 nm) increases significantly (from 3.5 ± 0.7% in 8G and 6.7 ± 0.5% in 12G to 15 ± 2% in 24G). This number is significantly higher than those calculated from simultaneous unfolding of the two noninteracting G-quadruplexes by chance (10.4 ± 0.2%, see Figure 2B). Such a result suggests the presence of quadruplex−quadruplex interactions in the 24G fragment. Given that the maximal coincident unfolding of two G-quadruplexes is 10.4% (see eq 2 for the calculation between 0 and 60 pN), we estimated a minimum of 4.6 ± 1.8% (15−10.4%) for the higher-order interactions between G-quadruplexes in the 24G telomeric fragment (Figure 4B).

To further clarify the presence of the higher-order interactions, we compared 2D histograms of rupture force ($F_{rupture}$) vs ΔL for 4G−24G fragments in Figure 5A. The region with high force (>30 pN) or high ΔL (>13 nm) clearly stands out in the 24G construct only (see the green ellipse in Figure 5A). Due to the additional stability imparted from the quadruplex−quadruplex interactions, it is expected that the rupture force of a higher-order structure is high. In addition, since the disruption of the tertiary interaction between quadruplexes likely leads to the cooperative unfolding of the interacting DNA secondary structures,[42,48] the ΔL is expected to be large. Therefore, the circled region in the 24G fragment can be ascribed to the higher-order interactions between G-quadruplexes. To quantify different regions statistically, we projected the 2D histograms onto the F−ΔL plane (Figure 5B) and analyzed with the HCA method on the 24G fragment (see Materials and Methods). At the 90% confidence level, four regions can be clearly distinguished in the 24G construct. On the basis of average ΔL values, the brown, red, and blue regions represent partially folded structure (ΔL = 5.7 ± 0.1 nm), G-quadruplex (ΔL = 8.7 ± 0.1 nm), and simultaneous unfolding of two noninteracting G-quadruplexes (ΔL = 16.4 ± 0.8 nm), respectively. The purported higher-order structures with either big ΔL or high rupture force (the green region) was clearly differentiated from the rest in the 24G construct. The percentage of this region is consistent with previous estimation for higher-order structures (∼5%), therefore, validating the assignments. The same HCA method failed to identify this higher-order region for other DNA constructs (data not

dx.doi.org/10.1021/ja510079u | J. Am. Chem. Soc. 2014, 136, 18062−18069

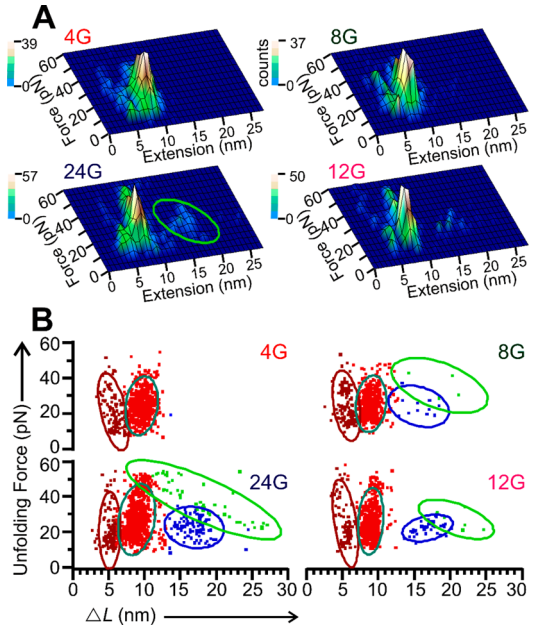


Figure 5. (A) 2D histograms of $F_{rupture}$ vs ΔL for the 4G, 8G, 12G, and 24G constructs. The green ellipse in the 24G construct suggests possible higher-order structures. (B) Projection of the 2D histograms from A onto the $F_{rupture}-\Delta L$ plane. In the 24G, four different populations were identified using the hierarchical cluster analysis (HCA), which were then used as a training set for the group classification (solid ellipses) of the 4G, 8G, and 12G constructs by the linear discriminant analysis (LDA) at 90% confidence level. Brown, cyan, blue, and green ellipses represent the partially folded structure, G-quadruplex, simultaneous unfolding of noninteracting G-quadruplexes, and higher-order G-quadruplexes, respectively. Except for the 24G construct, the higher-order structures are negligible in other constructs.

shown). For comparison, different regions in the 4G, 8G and 12G constructs were classified using LDA method based on the four classified regions determined in the 24G fragment (see Materials and Methods). In the 4G construct, the ΔL based LDA analysis only reveals partially folded and G-quadruplex structures. In the 8G and 12G constructs, LDA was able to identify all four populations as in the 24G. However, the regions for higher-order structures are negligible (0.6% and 0.7% of the total population for the 8G and 12G fragments respectively). This result is expected as G-quadruplexes follow beads-on-a-string model in the 8G and 12G constructs (see Figure 2).

Due to the ensemble average nature of biochemical and biophysical techniques, the small population of higher-order interactions (~5%) present in the 24G construct may not be easily observable in previous ensemble investigations. Since majority of the population (95%) does not show quadruplex-quadruplex interactions, we propose that the bead-on-a-string model is still valid for majority of G-quadruplexes in the 24G construct. Our observation, however, extends this model by revealing that there is a small portion of higher-order G-quadruplex structures in the long human telomeric overhang region. Based on the ΔL measurement in which simultaneous unfolding of 3 or more G-quadruplexes is negligible (Figure 4A), the higher-order structures most likely occur between two neighboring G-quadruplexes. Since side-to-side interaction of two G-quadruplexes can lead to the higher-order structures that are anchored by distal G-quadruplexes after looping out G-quadruplexes in the middle, which leads to more than two G-quadruplexes in the higher order structure, the long-range interaction may come from the G-quartet stacking between neighboring G-quadruplexes as suggested by literature.[26,49] Further support of stacking interaction between G-quadruplex units comes from mechanical unfolding of long telomeric RNA (TERRA).[50] RNA G-quadruplex is known to assume propeller structure that is prone to stack on each other.[27] Indeed, this is exactly observed in experiments.[51] When we mechanically unfolded the same 24G construct in a 10 mM Tris buffer (pH 7.4) with 100 mM NaCl, we found significantly reduced population of the larger ΔL population (8.4% for the 14.8 nm population, see Supporting Information Figure S7). Since this number is similar to that for simultaneous unfolding of two independent G-quadruplex (10.4% for ΔL 16.4 nm), we concluded that no high-order structure should exist in the Na^+ buffer. Previously, it has been determined that hybrid-1 and basket G-quadruplex are preferred conformations, respectively, at nanomolar concentrations using single-molecule methods.[29] Compared to the hybrid-1 conformation that has one propeller loop at the side of the G-quadruplex,[52] all three internal loops of the basket G-quadruplex stay at terminal G-quartet planes.[53] Due to the steric hindrance of these loops, the end-to-end stacking between G-quadruplex units is energetically costly, which explains the lack of the higher-order structures in the 24G construct in the sodium buffer. It is rather interesting that the quadruplex–quadruplex stacking does not go beyond two G-quadruplexes, presumably due to the large entropic penalty associated with the linearized arrangement of multiple stacking G-quadruplexes. When we evaluated the ΔG_{unfold} at 23 °C, a value of 33 ± 5 kcal/mol was obtained (see Materials and Methods and Supporting Information Figure S5 for calculation). This number is significantly higher than the ΔG_{unfold} of two noninteracting G-quadruplexes (~20 kcal/mol at 23 °C, see below for the calculation), which further validates our assignment of higher-order structures using the HCA method.

It is noteworthy that mechanical unfolding of protein oligomers, such as titins, also reveals a similar beads-on-a-string model for neighboring protein units.[54] Given the cooperative nature of protein structure in general, the stability of each immunoglobulin unit in titin is determined by intrinsic crosstalks between local conformations of different structural order. higher-order interaction such as domain swap between different protein units is therefore costly both entropically (due to long-range orderness) and enthalpically (due to disruption of original networking), which reduces the long-range interaction inside titin. Nucleic acid structures are known to be hierarchical with much simplified arrangement that does not involve significant interaction among local conformations of different structural order. This suggests long-range interaction is easier to occur due to reduced entropic penalty. The higher-order structures in the full-length telomeric overhang observed here for DNA and those for RNA[50] support this scenario.

The 24G Fragment Provides Single-Molecule Characterization of Telomeric G-Quadruplex with Increased Throughput. As the 24G fragment can host up to six G-quadruplexes, more quadruplex-unfolding events can be observed within one $F-X$ curve compared to the 4G fragment, in which only one G-quadruplex can form. Indeed, comparison of the percent formation between the 4G and the 24G reveals 3.8-fold increase in the G-quadruplex formation in the latter construct (Figure 6 A,B, see Materials and Methods for the estimation of the percent formation). Next, we compared the

dx.doi.org/10.1021/ja510079u | J. Am. Chem. Soc. 2014, 136, 18062–18069

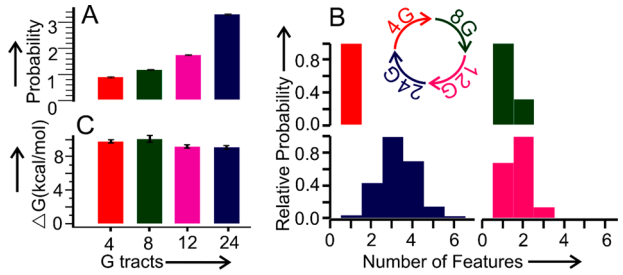


Figure 6. The 24G DNA construct offers increased throughput for mechanical unfolding experiments. (A) Probability of features observed per $(TTAGGG)_4$ for the four DNA constructs 4G, 8G, 12G, and 24G. (B) Probability of the number of features observed in each F–X curve for all four DNA constructs. (C) ΔG_{unfold} of noninteracting G-quadruplexes for all four DNA constructs.

properties of the G-quadruplexes formed in the two constructs. First, we deconvoluted the unfolding features in the 8G–24G constructs according to the regions classified by the HCA analysis in Figure 5B. After analyzing only the regions with noninteracting G-quadruplexes while excluding other structures (i.e., partially folded and higher-order structures, see Materials and Methods[31]), we found the changes in the free energy of unfolding a G-quadruplex (ΔG_{unfold}) were similar within experimental error in all four DNA constructs (Figure 6C, ΔG_{unfold} (23 °C) = 9.8 ± 0.2 to 9.1 ± 0.2 kcal/mol, see eq 5 for ΔG_{unfold} calculation). These ΔG_{unfold} values are similar to previously reported value of 9.8 ± 0.4 kcal/mol at 23 °C obtained from mechanical unfolding of the telomeric G-quadruplex.[43] They also fall into the range obtained from ensemble averaged methods (8.8 kcal/mol at 23 °C).[55] This fact, plus the identical ΔL and rupture force (see Figures 2 and 4) measured for noninteracting G-quadruplexes in the 4G–24G fragments, allowed us to conclude that noninteracting G-quadruplexes formed in these fragments are likely identical. Therefore, the 24G construct offers an increased throughput for mechanical unfolding experiments, which, due to its single-molecule nature, suffers from low efficiencies. It is interesting that the formation percentages of G-triplex are similar for all the DNA constructs (Supporting Information Figure S8). In addition, their mechanical stabilities are identical (Supporting Information Figure S9). Both observations suggest similar G-triplex conformations are formed in different G-rich constructs.

CONCLUSIONS

In this study, we have innovated a splint ligation strategy to prepare a telomeric DNA fragment with 24 repeating TTAGGG units (24G). Such a length falls within the range for the full-length 3′ human telomeric overhang. This method can be extended to prepare long DNA with multiple repeats of sequences that are difficult to synthesize. After investigating unfolding events in this 24G DNA construct, we have found a majority (95%) of the G-quadruplexes follow the bead-on-a-string model, while ∼5% partition into the higher-order quadruplex–quadruplex interactions. Compared to bulk experiments in which micromolar to millimolar template concentrations are used, the effective nanomolar concentration used here is closer to physiological relevant conditions. This fact may explain the controversial observation on the presence of the high order interaction between neighboring G-quadruplexes in ensemble experiments. As noninteracting G-quadruplexes in the 24G DNA construct are identical to those in the $(TTAGGG)_4$ construct, our method offers increased throughput for mechanical unfolding experiments at the single-molecular level.

ASSOCIATED CONTENT

Supporting Information

Methods and additional data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

hmao@kent.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to NSF CHE-1026532 and NSF CHE-1415883 for financial support. We are thankful to Cuixia Xu for critical reading of the paper.

REFERENCES

(1) Blackburn, E. H. *Nature* **1991**, *350*, 569.
(2) Levy, M. Z.; Allsopp, R. C.; Futcher, A. B.; Greider, C. W.; Harley, C. B. *J. Mol. Biol.* **1992**, *225*, 951.
(3) Chang, E.; Harley, C. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11190.
(4) Kim, N.; Piatyszek, M.; Prowse, K.; Harley, C.; West, M.; Ho, P.; Coviello, G.; Wright, W.; Weinrich, S.; Shay, J. *Science* **1994**, *266*, 2011.
(5) Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. *Genes Dev.* **1997**, *11*, 2801.
(6) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I., Jr.; Blackburn, E. H. *Cell* **1987**, *51*, 899.
(7) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Nature* **1991**, *350*, 718.
(8) Wang, Q.; Liu, J.-q.; Chen, Z.; Zheng, K.-w.; Chen, C.-y.; Hao, Y.-h.; Tan, Z. *Nucleic Acids Res.* **2011**, *39*, 6229.
(9) Lipps, H. J.; Rhodes, D. *Trends Cell Biol.* **2009**, *19*, 414.
(10) Yu, H.; Wang, X.; Fu, M.; Ren, J.; Qu, X. *Nucleic Acids Res.* **2008**, *36*, 5695.
(11) Zhao, C.; Wu, L.; Ren, J.; Xu, Y.; Qu, X. *J. Am. Chem. Soc.* **2013**, *135*, 18786.
(12) Hurley, L. H.; Wheelhouse, R. T.; Sun, D.; Kerwin, S. M.; Salazar, M.; Fedoroff, O. Y.; Han, F. X.; Han, H.; Izbicka, E.; Von Hoff, D. D. *Pharmacol. Ther.* **2000**, *85*, 141.
(13) Smith, F. F. W. J. *Nature* **1992**, *356*, 164.
(14) Kang, C.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* **1992**, *356*, 126.
(15) Yu, Z.; Schonhoft, J. D.; Dhakal, S.; Bajracharya, R.; Hegde, R.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2009**, *131*, 1876.
(16) Yu, H.; Gu, X.; Nakano, S.-i.; Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2012**, *134*, 20060.
(17) Adrian, M.; Ang, D. J.; Lech, C. J.; Heddi, B.; Nicolas, A.; Phan, A. T. *J. Am. Chem. Soc.* **2014**, *136*, 6297.
(18) Koirala, D.; Ghimire, C.; Bohrer, C.; Sannohe, Y.; Sugiyama, H.; Mao, H. *J. Am. Chem. Soc.* **2013**, *135*, 2235.
(19) Rajendran, A.; Endo, M.; Hidaka, K.; Sugiyama, H. *Angew. Chem., Int. Ed.* **2014**, *53*, 4107.
(20) Yue, D. J. E.; Lim, K. W.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 11462.
(21) Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2006**, *128*, 15461.
(22) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, 4649.

(23) Petraccone, L.; Spink, C.; Trent, J. O.; Garbett, N. C.; Mekmaysy, C. S.; Giancola, C.; Chaires, J. B. *J. Am. Chem. Soc.* **2011**, *133*, 20951.
(24) Petraccone, L. In *Quadruplex Nucleic Acids*; Chaires, J. B., Graves, D., Eds.; Springer: Berlin Heidelberg, 2013; Vol. *330*, p 23.
(25) Yatsunyk, L. A.; Piétrement, O.; Albrecht, D.; Tran, P. L. T.; Renčiuk, D.; Sugiyama, H.; Arbona, J.-M.; Aimé, J.-P.; Mergny, J.-L. *ACS Nano* **2013**, *7*, 5701.
(26) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876.
(27) Martadinata, H.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 2570.
(28) Collie, G. W.; Parkinson, G. N.; Neidle, S.; Rosu, F.; De Pauw, E.; Gabelica, V. *J. Am. Chem. Soc.* **2010**, *132*, 9328.
(29) Dhakal, S.; Cui, Y.; Koirala, D.; Ghimire, C.; Kushwaha, S.; Yu, Z.; Yangyuoru, P. M.; Mao, H. *Nucleic Acids Res.* **2013**, *41*, 3915.
(30) Luchette, P.; Abiy, N.; Mao, H. *Sens. Actuators, B* **2007**, *128*, 154.
(31) Dhakal, S.; Schonhoft, J. D.; Koirala, D.; Yu, Z.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2010**, *132*, 8991.
(32) Mao, H.; Luchette, P. *Sens. Actuators, B* **2008**, *129*, 764.
(33) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185.
(34) Yu, Z.; Mao, H. *Chem. Rec.* **2013**, *13*, 102.
(35) Schonhoft, J. D.; Bajracharya, R.; Dhakal, S.; Yu, Z.; Mao, H.; Basu, S. *Nucleic Acids Res.* **2009**, *37*, 3310.
(36) Yu, Z.; Cui, Y.; Selvam, S.; Ghimire, C.; Mao, H. *ChemPhysChem*, **2014**. DOI: 10.1002/cphc.201402443.
(37) Jurs, P. C.; B, G. A.; McClelland, H. E. *Chem. Rev.* **2000**, *100*, 2649.
(38) Formann, A. K. *Latent Class Analysis: Introduction to Theory and Application*; Beltz: Weinheim, 1984.
(39) Danielsson, P.-E. *Comput. Graphics Image Process.* **1980**, *14*, 227.
(40) Varshavsky, R.; Horn, D.; Linial, M. *PLoS One* **2008**, *3*, e2247.
(41) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(42) Yu, Z.; Gaerig, V.; Cui, Y.; Kang, H.; Gokhale, V.; Zhao, Y.; Hurley, L. H.; Mao, H. *J. Am. Chem. Soc.* **2012**, *134*, 5157.
(43) Koirala, D.; Dhakal, S.; Ashbridge, B.; Sannohe, Y.; Rodriguez, R.; Sugiyama, H.; Balasubramanian, S.; Mao, H. *Nat. Chem.* **2011**, *3*, 782.
(44) Koirala, D.; Mashimo, T.; Sannohe, Y.; Yu, Z.; Mao, H.; Sugiyama, H. *Chem. Commun.* **2012**, *48*, 2006.
(45) Laurence, T. A.; Kong, X.; Jager, M.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17348.
(46) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 2440.
(47) Moore, M. J.; Query, C. C. *Methods Enzymol.* **2000**, *317*, 109.
(48) Greenleaf, W. J.; Frieda, K. L.; Foster, D. A.; Woodside, M. T.; Block, S. M. *Science* **2008**, *319*, 630.
(49) Dai, J.; Carver, M.; Yang, D. *Biochimie* **2008**, *90*, 1172.
(50) Garavis, M.; Bocanegra, R.; Herrero-Galan, E.; Gonzalez, C.; Villasante, A.; Arias-Gonzalez, J. R. *Chem. Commun.* **2013**, *49*, 6397.
(51) Martadinata, H.; Phan, A. T. *Biochemistry* **2013**, *52*, 2176.
(52) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723.
(53) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263.
(54) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109.
(55) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic Acids Res.* **2008**, *36*, 5482.