J|A|C|S ARTICLES

Published on Web 01/20/2009

ILPR G-Quadruplexes Formed in Seconds Demonstrate High Mechanical Stabilities

Zhongbo Yu, Joseph D. Schonhoft, Soma Dhakal, Rabindra Bajracharya, Ravi Hegde, Soumitra Basu,* and Hanbin Mao*

Department of Chemistry, Kent State University, Kent, Ohio 44242

Received August 27, 2008; E-mail: sbasu@kent.edu; hmao@kent.edu

Abstract: The insulin linked polymorphism region (ILPR) is known to regulate transcription of the gene coding for insulin. The ILPR has guanine rich segments, suggesting that G quadruplexes may be responsible for this regulatory role. Using mechanical unfolding in a laser tweezers instrument and circular dichroism (CD) spectroscopy, we provide compelling evidence that highly stable parallel and antiparallel G quadruplex structures coexist in the predominant ILPR sequence of $(ACAGGGGTGTGGGG)_2$ at a physiologically relevant concentration of 100 mM KCl. Experiments at the single molecular level have shown that unfolding forces for parallel and antiparallel structures (F_{unfold}: 22.6 vs 36.9 pN, respectively) are higher than the stall forces of enzymes having helicase activities. From a mechanical perspective alone, these data support the hypothesis that G quadruplexes may cause replication slippage by blocking replication process. Using the unique combination of the rupture force and the contour length measured by laser tweezers, the simultaneous determination of probable parallel and antiparallel G quadruplex structures in a solution mixture has been achieved. Jarzynski's equality analysis has revealed that the antiparallel G quadruplex is thermodynamically more stable than the parallel conformer (ΔG_{unfold}: 23 vs 14 kcal/mol, respectively). On the other hand, kinetic measurements have indicated that both parallel and antiparallel structures fold rather rapidly (k_{fold}: 0.4 vs 0.3 s^{-1}, respectively), suggesting that they may be kinetically accessible for gene control. This work provides an unprecedented mechanical perspective on G quadruplex stability, presenting a unique opportunity to predict the functional consequence when motor enzymes encounter such structures.

Introduction

Unlike proteins or RNAs where secondary and higher order structures are common, DNA has a much lower propensity to form complex structures. A notable exception is the DNA G quadruplex where four single-stranded DNA (ssDNA) segments are joined by a stack of G quartets.[1,2] Each G quartet is composed of four coplanar guanines linked through Hoogsteen base pairs in a closed quadrilateral. Although the G quadruplex was first discovered in the 1960s[3], its physiological significance was not appreciated until quadruplexes were found in telomeres[4−8] and their presence *in vivo* was demonstrated.[9−11] A close scrutiny of the human genome sequences in recent years has shown widespread occurrence of G rich sequences that could form quadruplex structures.[12] Occurrence of these putative G quadruplexes is disproportionally higher close to or inside promoters of coding regions, implying possible regulatory roles for gene expression. Understanding such roles might depend on a detailed elucidation of specific G quadruplex structures since telomeric sequences have demonstrated a variety of topological formats.[13,14] However, structural information available for the G quadruplexes located outside the telomere region is sparse, hindering the detailed analysis of physiological roles of these quadruplex structures. Our paper will focus on the ILPR (Insulin Linked Polymorphism Region) located upstream of the promoter of the insulin gene and, therefore, may provide insightful perspectives to these rarely studied G quadruplexes.

Due to the limitation of available tools, current investigations often focus on intermolecular G quadruplex structures with slow kinetics. For example, although NMR can reveal detailed structural information,[15] it requires a significant amount of materials and is associated with low temporal resolution. Circular Dichroism (CD) has higher concentration sensitivity; however, it has difficulty following fast kinetic processes and

(1) Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703–30.
(2) Gilbert, D. E.; Feigon, J. *Curr. Opin. Struct. Biol.* **1999**, *9*, 305–14.
(3) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 2013–2018.
(4) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I., Jr.; Blackburn, E. H. *Cell* **1987**, *51*, 899–908.
(5) Oka, Y.; Thomas, C. A., Jr. *Nucleic Acids Res.* **1987**, *15*, 8877–8898.
(6) Sen, D.; Gilbert, W. *Nature* **1988**, *334*, 364–66.
(7) Sundquist, W. I.; Klug, A. *Nature* **1989**, *342*, 825–829.
(8) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871–880.
(9) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8572–8577.
(10) Duquette, M. L.; Handa, P.; Vincent, J. A.; Taylor, A. F.; Maizels, N. *Genes Dev.* **2004**, *18*, 1618–1629.
(11) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847–854.
(12) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908–16.
(13) Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14629–14634.
(14) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938–18943.
(15) Phan, A. T.; Kuryavyi, V.; Patel, D. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 288–298.

 10.1021/ja806782s CCC: $40.75

cannot always determine structures unambiguously,[16] especially when multiple structures are present in the same mixture. Other techniques, such as chromatography and gel electrophoresis,[17] can potentially perturb G quadruplex structures due to the interaction of quadruplexes with matrix. Since only a few identical copies of G-quadruplex forming DNA sequences exist *in vivo*, it is the intramolecular structure with relatively fast formation kinetics[18] that is likely to be physiologically relevant. Elucidating these intramolecular structures is an important first step to understanding their possible roles in gene regulation. Here, we will use single molecular methods to investigate intramolecular structures with relatively fast rates of quadruplex formation.

Single molecule methods provide novel platforms to investigate relatively fast kinetics of intramolecular structures that exist at rather low concentrations. Compared to bulk methods, single molecule approaches have the unique capability to examine individual subpopulations in a heterogeneous sample. This exceptional feature has been demonstrated by the observation of multiple G quadruplexes in recent single molecular FRET experiments.[13,14] Although highly sensitive in spatial and temporal measurement, the fluorescence method generally yields one-dimensional structural information and, therefore, could identify structural differences between parallel and antiparallel G quadruplexes only through very careful design. Moreover, the incorporation of relatively large fluorophores may perturb the structure under investigation.[18] Recently, optical tweezers based single molecule techniques have started to provide complementary structural information of DNA,[19,20] RNA,[21] and proteins[22] from an unprecedented perspective of mechanical stabilities. Since the method provides unique structural information through localized mechanical disturbance, instead of global perturbations in the thermal or chemical approaches, this method complements well with fluorescence approaches.

In addition, for a heterogeneous mixture of structures, the force method can simultaneously identify the presence of distinct species from the combined information of the contour length and the rupture force of individual structures. The determination of mechanical property of a biopolymer structure is of high physiological relevance, since biomacromolecules constantly experience various forces in the processes such as cell division or enzymatic reactions. For example, enzymatic action of motor proteins, such as DNA/RNA polymerases, can be stopped at a force known as the "stall force".

In this paper, we will use laser tweezers to investigate G quadruplexes in the ILPR.[23] This is the first time G quadruplex structures have been investigated through mechanical unfolding. Recent experiments have shown evidence consistent with *in vivo* formation of G quadruplexes,[24] which have been proposed to increase the expression of insulin.[25] Since insulin is directly related to the pathogenesis of diabetes, the potential regulatory role of the G quadruplex in the insulin expression has received close attention.[25–29] Compared to the extensively studied telomeric G quadruplex structures, however, there is a paucity of structural information on the G quadruplexes that form inside or close to promoter regions. We have found through our optical tweezers and CD experiments that both parallel and antiparallel G quadruplex structures form in the most prevalent ILPR repeat, $(ACAGGGGTGTGGGG)_2$. These structures are among the most stable human G quadruplexes characterized so far. In fact, their unfolding forces are greater than the stall forces of RNA polymerases[30] and enzymes with helicase activities,[31] suggesting from a mechanical perspective alone that they may endure the action of these enzymes in the replication or transcription processes.

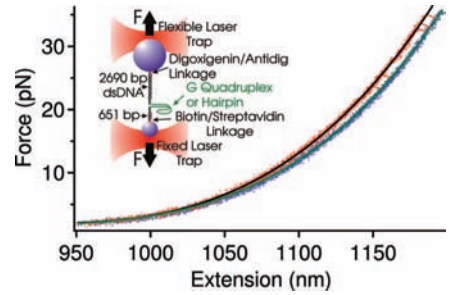

Figure 1. An overlap of five force extension curves of a 31-mer ILPR fragment in 100 mM KCl, 2 mM EDTA, 10 mM Tris buffer, pH 8.0. Loading rate was 5.5 pN/s. Red dots are for extending curves, while blue ones represent relaxing curves. Black and cyan lines are wormlike chain fits for extending and relaxing curves, respectively. Red lines between 20 and 35 pN indicate unfolding events. Inset shows a schematic of laser tweezers experiment.

Results and Discussion

Rupture Forces Show Two Populations. The predominant ILPR repeat has a sequence of 5′-ACAGGGGTGTGGGG[23]. When a DNA fragment based on this sequence, $5'$-$(ACAGGGGTGTGGGG)_2$, was inserted into a double-stranded DNA (dsDNA) construct and mechanically unfolded by laser tweezers (Figure 1 inset), we found the force extension curves showed unexpected sawteeth features, which are suggestive of complex structures such as intermolecular G quadruplexes. We reasoned that the dsDNA handle immediately downstream of the fragment may assume a bulky and stiff double helical conformation[20] that would interfere with the formation of intramolecular G quadruplexes. As a result, intermolecular structures might be favored. To reduce this steric hindrance, we added the wild type

(16) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
(17) Hardin, C. C.; Perry, A. G.; White, K. *Biopolymers* **2001**, *56*, 147–194.
(18) Mergny, J.-L.; Gros, J.; de Cian, A.; Bourdoncle, A.; Rosu, F.; Sacca, B.; Guittat, L.; Amrane, S.; Mills, M.; Alberti, P.; Takasugi, M.; Lacroix, L. *Energetics, Kinetics and Dynamics of Quadruplex folding. In Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; RSC Publishing: Cambridge, 2006; pp 31−80.
(19) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185–6190.
(20) Greenleaf, W. J.; Frieda, K. L.; Foster, D. A. N.; Woodside, M. T.; Block, S. M. *Science* **2008**, *319*, 630–633.
(21) Liphardt, J.; Onoa, B.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2001**, *292*, 733–737.
(22) Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S. *Science* **2005**, *309*, 2057–2060.
(23) Bell, G. I.; Selby, M. J.; Rutter, W. J. *Nature* **1982**, *295*, 31–35.
(24) Zhang, H.; Wang, Y.; Quinones, G.; Ligon, L. A.; McGown, L. B. *FASEB J.* **2008**, *22*, 778–3.
(25) Kennedy, G. C.; German, M. S.; Rutter, W. J. *Nat. Genet.* **1995**, *9*, 293–298.
(26) Lew, A.; Rutter, W. J.; Kennedy, G. C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12508–12512.
(27) Catasti, P.; Chen, X.; Moyzis, R. K.; Bradbury, E. M.; Gupta, G. *J. Mol. Biol.* **1996**, *264*, 534–545.
(28) Catasti, P.; Chen, X.; Mariappan, S. S. V.; Bradbury, E. M.; Gupta, G. *Genetica* **1999**, *106*, 15–36.
(29) Connor, A. C.; Frederick, K. A.; Morgan, E. J.; McGown, L. B. *J. Am. Chem. Soc.* **2006**, *128*, 4986–4991.
(30) Mejia, Y. X.; Mao, H.; Forde, N. R.; Bustamante, C. *J. Mol. Biol.* **2008**, *382*, 628–637.
(31) Perkins, T. T.; Li, H.-W.; Dalal, R. V.; Gelles, J.; Block, S. M. *Biophy. J.* **2004**, *86*, 1640–1648.

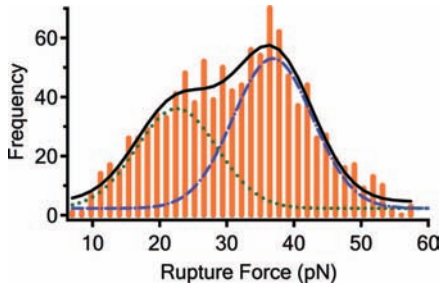

Figure 2. Histogram of rupture forces of the 31-mer ILPR sequence in 100 mM KCl, 2 mM EDTA, 10 mM Tris buffer, pH 8.0 ($n = 1069$). The entire population is fit by a two-peak Gaussian equation (solid line). The dotted line shows the Gaussian distribution of lower rupture force population; the dashed line shows the Gaussian distribution of higher force population.

ACA spacer sequence, which yielded the final DNA construct, 5′-(ACAGGGGTGTGGGG)$_2$*ACA*. Subsequent mechanical unfolding (Figure 1 inset) showed that sawteeth features had disappeared, and force extension curves (see Figure 1 for representative curves) were consistent with those of single molecules. The single molecule nature was further confirmed by a solitary breakage event when the tether was overstretched.

As a molecule was stretched by laser tweezers, the force experienced by the DNA construct increased until a rupture event occurred. Further stretching allowed the molecule to continuously extend along a trajectory characteristic of dsDNA handles (see Figure 1). This rupture event suggested the unfolding of a secondary structure. Analysis of 1069 curves revealed that the rupture force was rather spread out, from 7.7 pN to 58.5 pN with an average of 32.2 pN. Our current laser tweezers setup could not reach a force beyond 65 pN, at which one of the captured beads would escape from the trap. When all of the rupture forces were analyzed in a force histogram (Figure 2), to our surprise, two populations became obvious. A two-peak Gaussian fit yielded one population with a force centered at 22.6 pN (width 8.5 pN), while the other at 36.9 pN (width 8.4 pN).

After the force extension curves were fit by an extensible wormlike chain model,[19] change in contour length of the unfolded structure showed a relatively narrow distribution (see Supporting Information Figure S1, center of the distribution: 9.6 nm; Gaussian width: 1.7 nm. The distribution can also be fit by a two-peak Gaussian; see discussion below). This change in contour length was within the range calculated[32,33] for a 25-mer DNA fragment expected to form a G quadruplex, GGGGTGTGGGGACAGGGGTGTGGGG. Such a result strongly suggested that the rupture events were due to the unfolding of G quadruplex structures.

Two Populations Represent Parallel and Antiparallel G Quadruplex Structures. To determine the identities of the two populations in the rupture force histogram, CD experiments were performed. To match the fast refolding time in the single molecular experiments, which usually completes within a minute after unfolding a G quadruplex structure (see Materials and Methods), the CD signatures were collected after DNA fragments melted at 95 °C were quickly quenched to room temperature in ~30 s in an ice bath. In all of the CD experiments, we used 10 μM samples, which predominantly (~99%) formed intramolecular structures in native PAGE (data not shown).

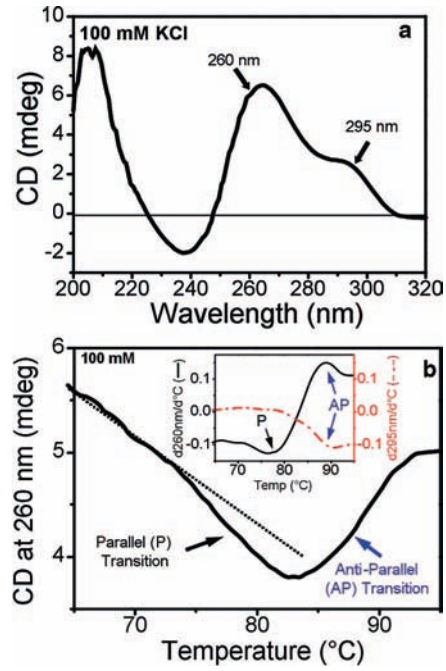

Figure 3. CD experiments of the 31-mer ILPR sequence in 10 mM Tris-HCl pH 8.0 with 100 mM KCl. (A) CD spectrum before melting. Arrows denote peaks at 260 and 295 nm that are characteristic of G-quadruplex structures (see text). Due to the significant temperature dependence of the 260 nm signal (see ref 48 and Figure 3B), CD peaks are not accurate for quantitation. (B) CD melting recorded at 260 nm. The temperature was ramped at 15 °C/h. Two transitions (~78 and ~88 °C) are marked by arrows. The dotted line is a guide for the baseline, which decreases linearly from 15 °C (not shown) to ~73 °C. The deviation from baseline that indicates the ~78 °C transition is real in seven different melting experiments (95% confidence level). Inset shows first derivative of CD signals at 260 nm (solid line, left axis) and 295 nm (dashed line, right axis) during melting. P and AP stand for parallel and antiparallel transitions, respectively.

The CD experiments on the same 31-mer DNA, (ACAGGGGTGTGGGG)$_2$ACA, clearly showed two peaks at ~260 and ~295 nm (Figure 3a). It is known that a peak at 260 nm and a trough at 240 nm correspond to a parallel G quadruplex structure, whereas a peak at 295 nm and a trough at 260 nm are associated with an antiparallel structure.[1] The observation of the two peaks at ~260 and ~295 nm is consistent with a mixture of parallel and antiparallel G quadruplexes or a hybrid G quadruplex consisting of both parallel and antiparallel segments.[34,35] To rule out the latter possibility, we have performed CD melting experiments and observed two major transitions at ~78 and ~88 °C (Figure 3b and inset). The first transition (~78 °C) was characterized by a decrease in the CD signal only at 260 nm, suggesting the melting of a parallel structure. On the other hand, the second transition (~88 °C) was accompanied by a signal increase at 260 nm and a decrease at 295 nm (Figure 3b inset). If this transition had occurred due to the melting of a hybrid G quadruplex, both 260 and 295 nm signals would have decreased (see Figure 8c in ref 34). Since this was not observed, we ascribed the second transition to the melting of an antiparallel structure. To confirm that Figure 3b

(32) Record, M. T. J.; Anderson, C. F.; Lohman, T. M. *Q. Rev. Biophys.* **1978**, *11*, 103–178.
(33) Mills, J. B.; Vacano, E.; Hagerman, P. J. *J. Mol. Biol.* **1999**, *285*, 245–257.
(34) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
(35) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. *J. Am. Chem. Soc.* **2006**, *128*, 1096–1098.

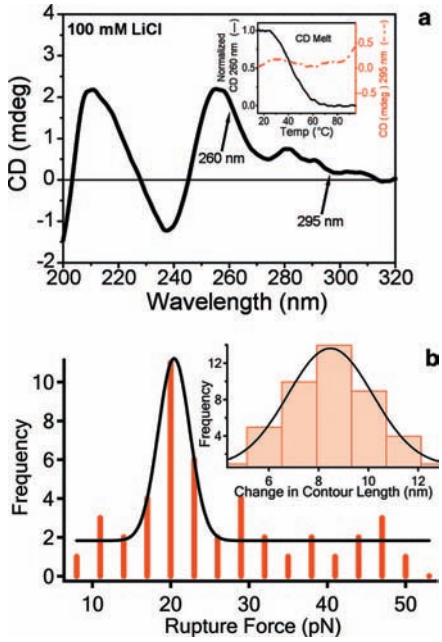

Figure 4. (A) CD spectrum of the 31-mer ILPR fragment in 10 mM Tris-HCl pH 8.0 with 100 mM LiCl before melting. The spectrum represents a mixture of a parallel G-quadruplex and nonstructured ssDNA (see text). Inset shows CD melting monitored at 260 nm (solid line, left axis) and 295 nm (dashed line, right axis). Only 260 nm shows a transition at ~45 °C. (B) Histograms of rupture force and change in contour length (inset) of the 31-mer ILPR fragment in the same buffer at 23 °C (n = 46). The solid lines are Gaussian fits.

represents the melting of parallel and antiparallel G quadruplexes, two ILPR variants that respectively form these two structures were mixed. CD melting of this mixture showed a similar pattern (see Figure S2 in Supporting Information). Due to the unfolding of the antiparallel structure, the negative trough at ~260 nm became less pronounced. This led to the apparent increase in the 260 nm signal, most of which was contributed by unstructured ssDNA as a result of quadruplex melting (data not shown).

These CD experiments implied that both parallel and antiparallel G quadruplexes could form within a minute in the 31-mer oligonucleotides. Further, they suggested that the antiparallel quadruplex was thermodynamically more stable than its parallel conformer. Therefore, a higher rupture force would probably be expected to unfold the antiparallel structures (Jarzynski's equality;[36,37] see discussion below).

To test whether the population exhibiting a higher rupture force belongs to the antiparallel structure, we repeated the experiments in the presence of Li^+, which has the least capability among the alkaline metals to stabilize G quadruplexes[38] and may demonstrate a preference for a specific quadruplex conformer. CD experiments in Li^+ showed that the presence of G quadruplex was substantially reduced while that of unstructured ssDNA significantly increased (Figure 4a). The existence of a parallel G quadruplex, instead of its antiparallel conformer, was revealed by the CD melting, where only one transition was shown at 260 nm and no obvious transition was observed at 295 nm (Figure 4a inset). The much reduced melting temperature, 42 °C for Li^+ compared to 78−88 °C for K^+ (Figure 3b), was consistent with the fact that lithium has a decreased capacity to stabilize G quadruplexes.

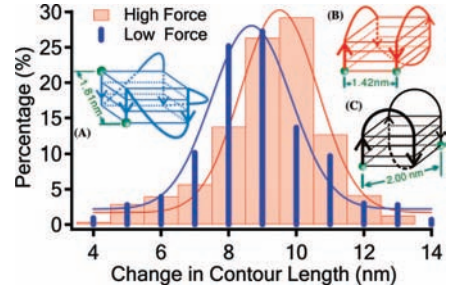

Figure 5. Histograms of change in contour length corresponding to lower (blue, n = 433, F < 29.8 pN) and higher (red, n = 636, F > 29.8 pN) rupture force populations. Structures corresponding to parallel (blue: molecule A) and antiparallel (red: molecule B; black: molecule C) G quadruplexes are shown. Two green balls in each structure represent two terminals where end-to-end distance is shown.

If the two populations in the rupture force reflected the unfolding of a parallel and an antiparallel G quadruplex in the presence of K^+ (Figures 2 and 3), then in accordance with the CD data in Li^+ (Figure 4a), we would expect only one rupture force population, namely, the parallel G quadruplex. This expectation was indeed observed (Figure 4b). Within experimental error, the force population in Li^+ was identical to the lower force population in K^+. The change in contour length (Figure 4b inset) was consistent with that of parallel G quadruplex (see discussion below). The likelihood of observing the folded G quadruplex in Li^+ at the single molecular level was one-fifth of that in K^+, reflecting the fact that lithium has a reduced capability to stabilize G quadruplexes. Based on these facts, we assign the lower force population to the parallel G quadruplex structure and the higher force population to the antiparallel strands. Our results clearly indicated the coexistence of both parallel and antiparallel ILPR quadruplexes in the same solution. Similar results were observed by single molecular FRET studies on telomere sequences.[13,14]

Rupture Force and Contour Length Determine Probable Structures for ILPR G Quadruplexes in Solution. After considering possible intramolecular G quadruplex structures,[16] we concluded that there was only one likely parallel form for the 31-mer ILPR sequence studied here (Figure 5, molecule A). On the other hand, two probable forms may contribute to CD signals characteristic of antiparallel strands (Figure 5, molecules B and C[34]). Although a hybrid G quadruplex structure consisting of both parallel and antiparallel strands may contribute to 260 and 295 nm peaks in CD signal, its presence was not consistent with our CD melting experiments, which indicated the existence of two species in solution (Figure 3b, see discussion above). Based on the results of Li^+ experiments (Figure 4), we concluded it was molecule A that was unfolded by low rupture forces. It is rather interesting that Li^+ prefers the parallel G quadruplex structure, which is a less stable conformer in K^+. We surmise that the small size of dehydrated Li^+ ions may prefer parallel G quadruplex structures whose G-quartet stacking is tighter than antiparallel conformers.[39,40] The structure that requires a higher rupture force will be decided by comparing the changes in contour length of B and C to that of A in the following paragraphs.

(36) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
(37) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2002**, *296*, 1832–1835.
(38) Venczel, E. A.; Sen, D. *Biochemistry* **1993**, *32*, 6220–6228.
(39) Wang, Y.; Patel, D. J. *J. Mol. Biol.* **1995**, *251*, 76–94.
(40) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880.

Based on the NMR[39] and X-ray crystal[40] structures of antiparallel and parallel telomeric G quadruplexes, respectively, we calculated that the end-to-end distance of structures A, B, and C are 1.81, 1.42, and 2.00 nm, respectively (see the distance between green balls in each molecule in Figure 5). Before a G quadruplex is unfolded, this end-to-end distance is an integral part of the apparent contour length obtained from the force extension curve. After the quadruplex is ruptured, however, the end-to-end distance and the overall contour length become identical for all three molecules. Therefore, the change in contour length, which is the difference between the contour lengths corresponding to the folded and unfolded portions of a force-extension curve, is decided by the end-to-end distance of unfolded G quadruplexes. Subsequent calculation revealed the difference of the change in contour length between A and B is 0.39 nm, whereas that between A and C is −0.19 nm.

Assuming that the lower and the higher force populations reflect the unfolding of parallel and antiparallel structures, respectively, we can identify their change in contour length by analyzing the two force populations simultaneously (see Materials and Methods for details). Figure 5 shows the two histograms of the change in contour length corresponding to lower and higher forces. The difference in the change of contour length between parallel and antiparallel structures were statistically significant at the 95% confidence level (*t* test, using an average of 9.3 nm and a standard deviation of 1.8 nm for parallel strands [$n = 433$]; and an average of 9.7 nm and a standard deviation of 1.6 nm [$n = 636$] for antiparallel strands). The variation of the change in contour length is 0.4 nm between parallel and antiparallel G quadruplex structures. Since this value is almost identical to the expected value between the A−B pair (0.39 nm), structure B most likely represents the antiparallel G quadruplex (as a comparison, the expected value between the A−C pair is −0.19 nm; see above). A similar antiparallel structure has been observed in an ILPR sequence by NMR.[27]

Within experimental error, the change in contour length of the G quadruplex in Li^+ (centered at 8.5 nm with a width of 2.4 nm, Figure 4b inset) was identical to that of the parallel G quadruplex in K^+ (centered at 8.7 nm with a width of 1.7 nm, Figure 5 blue histogram). This result strongly corroborated our assignment of the lower force population to the parallel G quadruplex.

Mechanical Unfolding of ILPR Quadruplexes Reveals an Exclusive View of Unexpectedly High Rupture Forces. When compared to the unfolding forces of DNA[20] and RNA[21] secondary structures or protein RNaseH,[22] which rarely exceed 25 pN, the rupture force for G quadruplex is unexpectedly high, especially for antiparallel strands (36.9 pN). Such a rupture force reflects the superb stability of G quadruplexes over the majority of biomacromolecules with known mechanical properties. Based upon mechanical forces alone, it can be argued that stable quadruplexes with a higher rupture force may affect the functionality of some motor enzymes with a lower stall force. For example, since the stall force of RNA polymerase is below 20 pN,[30] the unfolding force observed here suggests that ILPR G quadruplexes can withstand the action of RNA polymerase itself and block the transcription.

The higher rupture force for the antiparallel G quadruplex suggests that the antiparallel structure is probably more stable than the parallel quadruplex. This has been confirmed by nonequilibrium Jarzynski's equality analysis (see Materials and Methods).[36,37] While the antiparallel structure showed an unfolding free energy change (ΔG_{unfold}) of 23 ± 5 kcal/mol, the parallel structure had a ΔG_{unfold} of 14 ± 3 kcal/mol (see Materials and Methods). This large free energy difference has been corroborated by CD melting analysis (see Supporting Information). Previous studies have already demonstrated that, in an antiparallel ILPR G quadruplex, a guanosine in one of the lateral loops makes multiple stabilizing interactions with neighboring residues.[27] This could be a reason for the high stability observed in the antiparallel conformer.

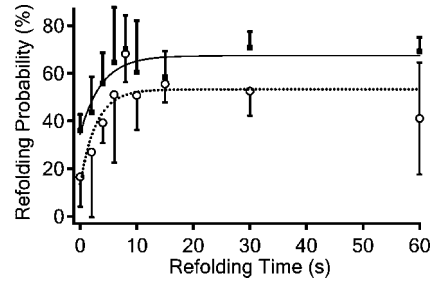

Figure 6. Refolding kinetics of the G quadruplexes formed in the 31-mer ILPR sequence. Experiments were performed in 100 mM KCl, 2 mM EDTA, 10 mM Tris buffer, pH 8.0. Empty circles with negative standard deviation (−sd; the last point shows ±sd) correspond to parallel G quadruplex, while filled squares with +sd correspond to antiparallel G quadruplex. Each data point is an average of at least three experiments. The data are fitted to a single exponential equation (see Materials and Methods). The dotted line is a fit for the parallel structure, while the solid line is a fit for the antiparallel structure.

Parallel and Antiparallel G Quadruplexes Form in Seconds. To investigate the refolding kinetics of G quadruplexes, we first mechanically unfolded a G quadruplex formed in the DNA construct. We then relaxed the force to 0 pN and incubated the sample for a specific time interval to allow the G quadruplex to refold. Subsequent force extension curves were recorded to reveal whether the G quadruplex refolded during the incubation. If the G quadruplex had refolded, a rupture force would have been observed. The type of the G quadruplex was determined by the magnitude of the rupture force (see Materials and Methods). By plotting the refolding probability of parallel and antiparallel structures separately with refolding time (Figure 6, details see Materials and Methods), we obtained the refolding rate constant for parallel (k_{fold}: 0.4 ± 0.2 s^{-1}) and antiparallel (k_{fold}: 0.3 ± 0.1 s^{-1}) G quadruplexes. These values implied that G quadruplexes could form within the lifetime of many cellular processes in the absence of other factors. The quantitative measurement of G quadruplex formation by single molecular experiments (within seconds) was within the range estimated by the qualitative CD method (within tens of seconds, Figure 3a).

Compared to the slow formation of G quadruplexes measured by bulk methods,[15,17] the fast folding kinetics measured by the single molecular method is rather surprising. The discrepancy between single molecular and bulk measurements probably reflects the difference in methodology. Unlike the facile characterization of both parallel and antiparallel G quadruplexes in the same solution as demonstrated here, the *ensemble* average nature of bulk experiments makes it extremely challenging to investigate individual subpopulations in a heterogeneous sample. Since *bona fide* single molecular conditions are difficult to obtain in bulk assays, the implication of slow intermolecular interactions[18] or complex equilibria between different subpopulations may cause the slower kinetics observed in bulk. The dsDNA handles flanking the G quadruplex used in our DNA construct impose natural constraints encountered by the *in vivo* formation of quadruplexes in G rich regions. Therefore, our system

represents physiological conditions more closely than other methods where these constraints are absent.[41]

Potential Physiological Roles of Parallel and Antiparallel G Quadruplexes. An *in vivo* G quadruplex may start to form once dsDNA containing a G rich sequence is unwound, which occurs during replication or inside a transcription bubble. Our recent results have demonstrated that, in the presence of a complementary strand, the formation of a G quadruplex in the ILPR sequence is kinetically competitive but thermodynamically prevailing compared to the formation of dsDNA (data not shown). The G quadruplex is proposed to cause the replication slippage that leads to variable number tandem repeats (VNTR) by serving as a block to replication.[27,28] Replication usually starts with the unwinding of dsDNA by enzymes with helicase activities. Our finding that the unfolding forces of G quadruplexes are higher than the stall force of enzymes with helicase activities (~20 pN or lower[31]) supports the notion that a stable quadruplex structure can mechanically block these enzymes. However, it is not yet known whether the unwinding capacity of an enzyme depends on the stability of a particular G quadruplex structure. Based on the fact that DNA helicase cannot unwind a G quadruplex stabilized by a ligand,[42,43] it is likely that the enzyme may show different activities toward parallel or antiparallel G quadruplex conformers that vary in thermodynamic or mechanical stabilities. Replication slippage could therefore be potentially regulated through the interconversion between parallel and antiparallel G quadruplexes, which may be accomplished in the context of complementary C rich strands, ILPR variants with different sequences, and various helicases or other interacting proteins.

Conclusions

We have, for the first time, applied mechanical unfolding to investigate the G quadruplexes formed in the ILPR region. Highly stable parallel and antiparallel G quadruplexes have been found to coexist under physiologically relevant conditions, with antiparallel as more stable than parallel conformers. Unfolding forces for these structures are higher than the stall forces of enzymes having helicase activities, supporting the hypothesis that G quadruplexes could cause replication slippage by blocking the replication process from a mechanical perspective alone. The rapid formation of both G quadruplexes suggests their kinetic accessibility for control of gene expression. Assisted with complementary techniques such as CD spectroscopy, the capability of simultaneous structural determination of probable parallel and antiparallel G quadruplexes in a solution mixture further demonstrates the unique strength of this method.

Materials and Methods

Laser Tweezers Instrument. Detailed description of the laser tweezers instrument has been reported elsewhere.[44,45] Briefly, a diode pumped solid state (DPSS) laser (1064 nm, 4 W, CW mode, BL-106C, Spectra-physics) was used as a trapping laser. *P* and *S* polarized laser light from the same laser source constituted two traps. The *S* polarized light was controlled by a steerable mirror (Nano-MTA, Mad City Laboratories) at a conjugate plane of the back focal plane of a focusing objective (Nikon CFI-Plan-Apochromat 60×, NA 1.2, water immersion, working distance ~320 μm). The exiting *P* and *S* polarized beams were collected by an identical objective and detected by two position sensitive photodetectors (PSD, DL100, Pacific Silicon Sensor)[46] separately.

The force of the laser trap was calibrated by the Stokes force and thermal motion measurement. Both methods yielded a similar trap stiffness of ~307 pN/(μm × 100mW) (for 0.97 μm diameter polystyrene beads, Bangs Laboratory, Fishers, IN). To test the accuracy of this instrument in the force measurement, we have unfolded a known RNA hairpin structure[47] tethered between two trapped polystyrene beads through DNA−RNA hybrid handles (samples were kindly provided by Dr. Wei Cheng at UC Berkeley). The hairpin was mechanically unfolded (similar to Figure 1 inset, unfolding details see below) at ~23 pN, a value consistent with the literature.[47]

DNA Construct. To prepare the DNA construct containing the 31-mer ILPR (Figure 1, inset), PCR amplification of pEIB plasmid (966bp)[46] was carried out in the presence of a biotinylated primer (Integrated DNA Technologies, IDT, Coralville, IA). Purified (Qiagen, Germantown, MD) DNA was then digested overnight with *XbaI* (NEB, Ipswich, MA).

Two oligonucleotides purchased from IDT (5′-CGCATCTGTGCGGTATTTCACACCGT, and 5′-GGCCGACGCGCTGGGCTACGTCTTGCTGGC) were annealed with a third oligonucleotide that contains a 31-mer ILPR sequence (underlined) (5′-CTAGACGGTGTGAAATACCGCACAGATGCG**ACAGGGGTGTGGGGACAGGGGTGTGGGGACA**GCCAGCAAGACGTAGCCCAGCGCGTC) at 97 °C, followed by slow cooling to room temperature for 6 h. The DNA piece so constructed was phosphorylated at the 5′ end by a polynucleotide kinase (NEB).

By using T4 DNA ligase (NEB), DNA fragments of 651 bp and 315 bp from *XbaI* digestion of the pEIB plasmid were ligated with the DNA piece constructed above. The ligation product derived from the 651 bp fragment was purified by a gel extraction kit (Midsci, St. Louis, MO).

SacI (NEB) linearized pEGFP plasmid DNA (Clontech, Mountain View, CA) was ethanol precipitated and digested with *EagI* (NEB). The longest DNA fragment (2690bp) was gel purified. It was labeled by digoxigenin (Dig) at the 3′ end by terminal transferase (Fermentas, Glen Burnie, MD) using 18 μM Dig-dUTP (Roche, Indianapolis, IN). Excess Dig-dUTP (Dig-dUTP: 2690bp DNA = 150:1 nmol) was used to ensure the complete Dig labeling. The Dig labeled 2690bp DNA fragment was ethanol precipitated and eluted in 10 mM Tris (pH 8.0). Dig labeling was confirmed by 3 h of incubation with anti-Dig antibody, which showed a smear of bands with slower mobility after agarose gel electrophoresis.

Final ligation between the Dig labeled 2690bp and DNA fragment containing ILPR was conducted in a special ligation solution consisting of 30% DMSO and 8% PEG-6000 (Sigma). To minimize the dimerization of Dig-2690bp, it was added in four aliquots every 2 h, with a final ratio of 1:3 between the Dig-2690bp and the DNA fragment containing ILPR. Ligation was performed at 37 °C for 8 h followed by 0.5 h cycling between 30 and 22 °C for another 8 h. Final purification was conducted by phenol-chloroform extraction followed by ethanol precipitation. The resulting pellet was dissolved in 10 mM Tris (pH 8.0, concentration 8.3 ng/μL) and stored at −20 °C.

Single Molecular Experiments. Streptavidin-coated polystyrene beads (diameter: 0.97 μm, Bangs Laboratory) were incubated with the diluted Dig labeled DNA construct obtained above (~0.166 ng/μL) in 100 mM KCl, 2 mM EDTA, 10 mM Tris buffer, pH 8.0) for 1 h at 23 °C before they were trapped by laser tweezers.

(41) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic Acids Res.* **2008**, *36*, 5482–5515.
(42) Han, H.; Bennett, R.; Hurley, L. *Biochemistry* **2000**, *39*, 9311–9316.
(43) Wu, X.; Maizels, N. *Nucleic Acids Res.* **2001**, *29*, 1765–1771.
(44) Luchette, P.; Abiy, N.; Mao, H. *Sens. Actuators, B* **2007**, *128*, 154–160.
(45) Mao, H.; Luchette, P. *Sens. Actuators, B* **2008**, *129*, 764–771.
(46) Mao, H.; Arias-Gonzalez, J. R.; Smith, B.; Tinoco, I., Jr.; Bustamante, C. *Biophys. J.* **2005**, *89*, 1308–1316.
(47) Dumont, S.; Cheng, W.; Serebrov, V.; Beran, R. K.; Tinoco, I., Jr.; Pyle, A. M.; Bustamante, C. *Nature* **2006**, *439*, 105–108.
(48) Petraccone, L.; Erra, E.; Esposito, V.; Randazzo, A.; Galeone, A.; Barone, G.; Giancola, C. *Biopolymers* **2005**, *77*, 75–85.

Anti-Dig antibody coated beads (diameter: 2.17 μm, Spherotech, Lake Forest, IL) were dispersed into the same buffer before use.

These two types of beads were captured separately by the two laser traps. Since one end of the DNA construct was modified with biotin and the other with Dig, the DNA construct could be immobilized between the two beads *via* the streptavidin/biotin and Dig/anti-Dig complexes, respectively. To this end, the laser trapped streptavidin-coated polystyrene bead bound with DNA constructs was approached by the anti-Dig coated bead until a tether between the two beads was observed (Figure 1, inset). The anti-Dig coated bead, controlled by the Nano-MTA positioner, was then moved away from the streptavidin-coated bead with a loading speed of ~5.5 pN/s. During this process, the tension inside the molecule increased until any secondary structure was mechanically unfolded (Figure 1). The presence of a single tether was confirmed by the observation of a single breakage event when the tether was overstretched after each experiment.

Rupture events with changes in contour length between 5 and 14 nm were included in the histograms of rupture force and change in contour length (Figures 2, S1, and 4b). To measure the refolding kinetics, G quadruplexes were first mechanically unfolded. They were then relaxed to 0 pN for a specific time (0, 2, 4, 6, 8, 10, 15, 30, and 60 s) prior to subsequent pulling, which was used to indicate whether G quadruplexes had been refolded during the incubation. The refolding probability was determined by the number of folded events divided by the total events. The rupture forces were grouped into low (<29.8 pN) and high (>29.8 pN) force regions separately. The refolding probability (y) at a specific incubation time (t) was calculated for these two force regions. The refolding rate constant (k_{fold}) was obtained by a single exponential fit: $y = y_0 + A \exp(-k_{fold}t)$ (Figure 6), where y_0 and A are constants.

The contour length (L) of the secondary structure was obtained through the fitting of a force extension curve to an extensible WLC equation:[19]

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

where x is the end-to-end distance, F is force, S is the elastic stretch modulus (1226 pN[19]), P is the persistent length (51.95 nm[19]), k_B is the Boltzmann constant, and T is absolute temperature.

Calculation of ΔG_{unfold}. The unfolding free energy difference (ΔG_{unfold}) for parallel and antiparallel G quadruplexes was calculated according to Jarzynski's equality equation:[36,37]

$$\Delta G = -k_B T \ln \sum_{i=1}^{N} \frac{1}{N} \exp\left(-\frac{W_i}{k_B T}\right)$$

where N is the number of repetitions in the experiment and W is the nonequilibrium work done to unfold the G quadruplexes, which equals the hysteresis area between stretching and relaxing force extension curves.[20] A total of 55 and 41 curves were calculated to represent the free energy difference of antiparallel and parallel G quadruplexes, respectively. These two sets of curves had a Gaussian distribution pattern identical to that described in Figure 2.

CD Spectroscopy. A 10 μM solution of the 31-mer ILPR DNA (IDT) was prepared in 10 mM Tris-HCl buffer (pH 8.0) with 100 mM potassium chloride or 100 mM lithium chloride in a total volume of 200 μL. To prevent cross contamination of ions, samples were ethanol precipitated three times before dissolving in a buffer with a specific ion. Before measurement, the DNA samples were heated at 97 °C for 10 min and cooled to room temperature within 30 s using an ice−water bath. CD experiments were performed using a Jasco-810 spectropolarimeter equipped with a Jasco (model PFD-425S) peltier temperature controller (Easton, MD). All experiments were performed using a quartz cuvette with a 1 mm optical path length with ~100 μL of mineral oil added on top of the solution to prevent evaporation. The spectra obtained were the average of three scans at a 0.5 nm interval with a scan rate of 50 nm/min. The resulting spectra were then subtracted from a buffer-only baseline and smoothed using a Savitzky−Golay function. For melting experiments, the CD value at 260 or 295 nm was recorded at 0.5 °C intervals while the temperature was increased at a rate of 15 °C/h. First derivatives of CD melting curves were obtained in Origin 6.1 software by using the Savitzky−Golay smoothing/derivative function.

Acknowledgment. H.M. thanks a New Faculty Award Program at Camille and Henry Dreyfus Foundation, a KSU startup, and Ohio Board of Regents for financial support. S.B. acknowledges a KSU startup fund and an Ohio Board of Regents fund. We thank Dr. Frederick Walz for the critical reading of this manuscript.

Supporting Information Available: Histogram and CD melting plots; calculation details for free energy change from CD melting experiments and for unfolding kinetics of G quadruplexes. This material is available free of charge via the Internet at http://pubs.acs.org.

JA806782S