11044–11056 *Nucleic Acids Research, 2019, Vol. 47, No. 21* *Published online 16 October 2019*
doi: 10.1093/nar/gkz906

Streamlining effects of extra telomeric repeat on telomeric DNA folding revealed by fluorescence-force spectroscopy

Jaba Mitra[1,2] and Taekjip Ha[2,3,4,5,*]

[1]Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana IL 61801, USA, [2]Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, Baltimore, MD 21205, USA, [3]Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218, USA, [4]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA and [5]Howard Hughes Medical Institute, Johns Hopkins University, Baltimore, MD 21218, USA

Received June 18, 2019; Revised September 29, 2019; Editorial Decision September 30, 2019; Accepted October 02, 2019

ABSTRACT

A human telomere ends in a single-stranded 3′ tail, composed of repeats of T_2AG_3. G-quadruplexes (GQs) formed from four consecutive repeats have been shown to possess high-structural and mechanical diversity. In principle, a GQ can form from any four repeats that are not necessarily consecutive. To understand the dynamics of GQs with positional multiplicity, we studied five and six repeats human telomeric sequence using a combination of single molecule FRET and optical tweezers. Our results suggest preferential formation of GQs at the 3′ end both in K^+ and Na^+ solutions, with minor populations of 5′-GQ or long-loop GQs. A vectorial folding assay which mimics the directional nature of telomere extension showed that the 3′ preference holds even when folding is allowed to begin from the 5′ side. In 100 mM K^+, the unassociated T_2AG_3 segment has a streamlining effect in that one or two mechanically distinct species was observed at a single position instead of six or more observed without an unassociated repeat. We did not observe such streamlining effect in 100 mM Na^+. Location of GQ and reduction in conformational diversity in the presence of extra repeats have implications in telomerase inhibition, T-loop formation and telomere end protection.

INTRODUCTION

As early as 1962, guanosine moieties were known to self-assemble into tetrameric structures via Hoogsteen hydrogen bonds (1). Planar association between four guanines by eight hydrogen bonds in the presence of a central coordinating monovalent cation (e.g. Na^+ and K^+) generates a G-quartet or a G-tetrad (2,3). Stacking of two or more G-quartets creates a very stable DNA secondary structure called the G-Quadruplex (GQ) (4). Complete sequencing of the human genome has identified about 300 000 putative sequences that can fold into GQs (5). Expansion of GQ-forming motifs have been implicated in pathogenicity associated with several human neurological disorders (6). Additionally, GQ-forming sequences have been reported in various viral genomes including human immuno-deficiency virus, Epstein–Barr virus and human papillomavirus (7).

In human cells, telomeric DNA continues beyond the double-stranded region as a 3′ 100–200 nucleotides (nt) single-stranded DNA made of repeats of hexanucleotide T_2AG_3 (8). This G-rich overhang of T_2AG_3 repeats form GQ structures under physiological conditions (4,9–10). Telomere length homeostasis has important implications in cell survival and proliferation (11–14) and telomeric GQs are studied as targets for cancer treatment and for potential applications in nanotechnology (15–17).

Despite an apparently simple, repeating G-rich sequence, structural studies with circular dichroism (CD), nuclear magnetic resonance and X-ray crystallography have revealed extreme polymorphism in human telomeric GQs composed of four T_2AG_3 repeats (18–24). Such diversity suggests co-existence of different conformations in the telomere with implications on rational designing of GQ-targeting drugs. GQ polymorphism and conformational dynamics in longer telomeric DNA containing five or more T_2AG_3 repeats, which should better mimic a telomere overhang, have been examined only in a few studies (25–29).

GQ formation requires a minimum of four G-rich segments (30). As human telomeric DNA contains multiple hexanucleotide (T_2AG_3) repeats and stretches for several kilobases, it has been assumed that GQs can form anywhere along the G-rich strand (31). Hence, under physiological conditions, a long telomeric DNA with more than

*To whom correspondence should be addressed. Tel: +1 410 614 4039; Fax: +1 410 955 0637; Email: tjha@jhu.edu

four T_2AG_3 repeats can harbor different forms of terminal structures depending on the position of GQ(s). GQ locations on telomeric overhangs can modulate telomerase extension or mechanism of alternative lengthening of telomere (ALT) (14). An exonuclease hydrolysis assay suggested preferential GQ formation via association of four consecutive hexanucleotide repeats at the 3′ end of long telomeric DNA, leaving out a T_2AG_3 repeat at the 5′ end (26). Another study employing site-specific labeling of guanine residues suggested the presence of hybrid GQ structures, incorporating one or more hexanucleotide repeat within a single loop to connect adjacent Gs in a G-quartet (29). Ensemble average techniques used in the above studies are limited by time resolution and hence cannot univocally deconvolute individual species from a heterogeneous population. Dynamic exchanges between GQ conformations in telomeric DNA spanning five to seven T_2AG_3 repeats have been observed using single molecule fluorescence resonance energy transfer (smFRET) (32). In addition, single molecule optical tweezers studies of human telomeric DNA of four to seven repeats indicated that GQs more frequently form using four consecutive repeats (25). In a recent study we demonstrated extreme mechanical diversity of GQs in 22 nt long human telomeric DNA (four telomeric repeats), where at least six different mechanically distinct species were observed (30). However, despite the physiological relevance, diversity in mechanical responses of GQs formed in longer telomeric DNA is yet to be explored.

In this study, we employed smFRET to probe GQ formation of human telomeric DNA, spanning five and six T_2AG_3 repeats. Using guanine to thymine (G to T) mutation at the central G of a T_2AG_3 repeat, which is known to restrict the participation of the repeat in GQ formation (24,30), we created variants without positional multiplicity to help interpret the data. Using smFRET and its combination with optical tweezers, which we refer to as fluorescence-force spectroscopy, we found evidence for preferential formation of GQs at the 3′ end of five-repeat telomeric DNA. GQs formed at the 5′ end or at an internal position were minority populations. GQs formed at the 3′ end unravel non-cooperatively via strand slippage (discussed later) in contrast to the cooperative unfolding of canonical GQs formed from four telomeric repeats (30). Furthermore, using a superhelicase-based vectorial folding assay (33–35), we show that a GQ forms preferentially at the 3′ end even when it is allowed to fold starting from the 5′ side. Although constituting a minor population, GQs formed at the 5′ end of a five-repeat telomeric DNA show extreme mechanical stability and cannot be perturbed by forces up to ∼28 pN. GQs preferentially formed at the 3′ end of six telomeric repeats and manifested similar mechanical properties. Overall, we found that an unassociated T_2AG_3 repeat has a streamlining effect on GQ conformations in that one or two mechanically distinct conformations are observed at a single location instead of six or more observed for GQs without an unassociated repeat. Therefore, we should consider the possibility the telomeric DNA and G-rich DNA in general may have much simpler folding landscapes than what have deduced from the minimal GQ-forming sequences.

MATERIALS AND METHODS

DNA constructs

All DNA oligonucleotides were purchased from Integrated DNA Technologies. The telomeric constructs were designed in lines with that used in a previous study by Mitra *et al.* (30). The five and six telomeric repeat sequences referred to as hTel28 and hTel34, respectively, are as follows:

5′-TGGCGACGGCAGCGAGGC**GGG(TTAGGG)$_4$T**/Cy3/T_{17}TCGGGAGCGGACGCACGG-3′ (hTel28) and 5′-TGGCGACGGCAGCGAGGC**GGG(TTAGGG)$_5$T**/Cy3/T_{17}TCGGGAGCGGACGCACGG-3′ (hTel34), where the telomeric repeat motif is bold-faced. For the repeat site selective mutation studies, five hTel28 mutants were designed such as GTG(TTAGGG)$_4$T (hTel28.1), GGG(TTAGTG)(TTAGGG)$_3$T (hTel28.2), GGG(TTAGGG)(TTAGTG)(TTAGGG)$_2$T (hTel28.3), GGG(TTAGGG)$_2$(TTAGTG)(TTAGGG)T (hTel28.4) and GGG(TTAGGG)$_3$(TTAGTG)T (hTel28.5), where the underline denotes mutation form G to T. A truncated GQ strand hTel16, where GGG(TTAGGG)$_3$T was replaced with GGG(TTAGGG)$_2$T, was also used in our experiments. For further probing GQ formation in hTel34, an internal position was labeled with Cy3 and referred to as hTel34′: 5′-TGGCGACGGCAGCGAGGC**GGG(TTAGGG)$_3$T**/Cy3/**(TTAGGG)$_2$TT**$_{17}$TCGGGAGCGGACGCACGG-3′. Another sequence designed to the same effect was designated as 6T.hTel22.6T: 5′TGGCGACGGCAGCGAGGC**T$_6$GGG(TTAGGG)$_4$T$_6$**/Cy3/T_{17}TCGGGAGCGGACGCACGG-3′. A complementary stem strand of sequence 3′/Biotin/ACCGCTGCCGTCGCTCCG/Cy5/5′ was used to immobilize the telomeric sequences of interest onto the slide surface. A second complementary strand (λ-bridge) of sequence 3′ AGCCCTCGCCTGCGTGCCTCCAGCGGCGGG 5′ was used to bridge the hTel strand with the 12 nt COS site of the λ-DNA. The other end of the λ-DNA could be further annealed with a digoxigenin labeled strand of sequence 3′AGGTCGCCGCCC/dig/5′. The hTel strands were first annealed with the complementary stem in 1.1:1 ratio in a buffer containing 50 mM NaCl and 10 mM Tris–HCl, pH 8, at 95°C for 5 min, followed by slow cooling to room temperature. The final construct for single molecule experiments by total internal reflection microscopy (TIRF) was generated by adding the λ-bridge DNA to the above mixture in the ratio of 1.5:1 to the complementary stem strand followed by incubation with rotation at room temperature for an hour.

For integrated smFRET-optical tweezers assay, λ-DNA (16 nM, New England Biolabs) was first heated in presence of 120 mM Na^+ at 80°C for 10 min and then quenched on ice for 5 min. The hTel constructs and bovine serum albumin (BSA) were added to the λ-DNA at a final concentration of 8 nM and 0.1 mg/ml, respectively, and incubated/rotated at room temperature for 2–3 h. The dig-strand was next added to a final concentration of 200 nM and then incubated with rotation at room temperature for 1 h.

Sample assembly

In order to eliminate non-specific surface binding, coverslips and quartz slides/glass slides used in our experiments were passivated with polyethylene-glycol (PEG) (a mixture of mPEG-SVA and biotin-PEG-SVA, Laysan Bio) (36). The PEGylated surfaces of the slides and coverslips were further sandwiched to form imaging chambers. For TIRF experiments, ∼30 pM hTel construct(s) were immobilized on the PEGylated surface via biotin-neutravidin interaction and imaged in a buffer containing 20 mM Tris–HCl pH 8, 0.8% w/v D-Glucose [Sigma], 165 U/ml glucose oxidase [Sigma], 2170 U/ml catalase [Roche], 3 mM Trolox [Sigma] and pre-determined amount of NaCl or KCl.

For integrated smFRET-optical tweezers experiments, a similar imaging chamber was incubated in blocking buffer containing 10 mM Tris HCl pH 8, 50 mM NaCl, 1 mg/ml BSA [NEB] and 1 mg/ml tRNA [Ambion] for 1 h. The DNA constructs were immobilized on the surface at ∼10 pM concentration, via biotin-neutravidin interaction. Next, anti-digoxigenin coated polystyrene beads (Polysciences) were attached to the immobilized construct by incubation in a buffer containing 10 mM Tris HCl pH 8 and 50 mM NaCl for 30 min. At last, data were acquired in an imaging buffer consisting of 50 mM Tris–HCl pH 8, 0.8% w/v D-Glucose [Sigma], 0.5 mg/ml BSA [NEB], 165 U/ml glucose oxidase [Sigma], 2170 U/ml catalase [Roche], 3 mM Trolox [Sigma] and pre-determined amount of NaCl/KCl.

Fluorescence-force spectroscopy

An integrated single molecule fluorescence-optical trap instrument was recently developed in our lab to probe conformational changes of biomolecular systems under tension (37,38). In short, an optical trap was formed by an infrared laser (1064 nm, 800 mW (maximum average power), EXLSR-1064-800-CDRH, Spectra-Physics) through the back port of the microscope (Olympus) on the sample plane with a 100× immersion objective (Olympus). Tension was applied on the sample tethers by translating the piezoelectric stage that holds the microscope and the applied force was read out via position detection of the tethered beads using a quadrant photodiode (UDT/SPOT/9DMI). The instrument was calibrated as described previously (37,38). A confocal excitation laser (532 nm, 30 mW (maximum average power), World StarTech) was used to scan the sample via a piezo-controlled steering mirror (S-334K.2SL, Physik Instrument). Two avalanche photodiodes were used to record the fluorescence emission, filtered from infrared laser by a band pass filter (HQ580/60 m, Chroma) and excitation by a dichroic mirror (HQ680/60 m, Chroma.

Data acquisition

In the absence of force, a prism-type total internal reflection microscope, with 532 nm laser excitation and back-illuminated electron-multiplying charge-coupled device camera (iXON, Andor Technology, South Windsor, CT, USA) was used to detect smFRET signals (36). The donor and acceptor intensities, i.e. I_D and I_A, respectively, were corrected for background signals and crosstalk and used to estimate smFRET efficiency E using $I_A/(I_A+I_D)$. E histograms were constructed by averaging the first ten data points of each molecule's time trace.

A detailed data acquisition procedure for integrated single molecule force-fluorescence spectroscopy has been described by Hohng *et al.* (37). Briefly, a tethered bead was trapped and its origin was determined by stretching the tether along opposite directions along x- and y-axes. The trapped bead was next moved from its origin by 14 μm and fluorescence spot on the tether located by scanning with the confocal. All unfolding and/or subsequent refolding experiments were performed by translating the microscope stage at a speed of 455 nm/s between 14 μm and 16.8–17.2 μm. Fluorescence emission from the tethered molecule was detected simultaneously with the stage movement, 20 ms after each step in the stage movement. Similar to TIRF-based smFRET experiments, force-fluorescence data were obtained in the imaging buffer containing 50 mM Tris–HCl pH 8, 0.8% w/v D-Glucose [Sigma], 0.5 mg/ml BSA [NEB], 165 U/ml glucose oxidase [Sigma], 217 U/ml catalase [Roche], 3 mM Trolox [Sigma] and pre-determined amount of NaCl/KCl.

The f_{unfold} histograms were analyzed and the free energy parameters such as the transition distances to unfolding etc were predicted with reference to the Dudko–Szabo model ($\nu = 1/2$). These derived parameters ($\Delta x^{\ddagger}$, $\tau_u(0)$, and $\Delta G^{\ddagger}$) were then used to reconstruct the force profiles (39,40).

Circular dichroism (CD) spectroscopy

CD spectra of the human telomeric oligonucleotides were recorded on an Aviv-420 spectropolarimeter (Lakewood, NJ, USA), using a quartz cell of 1 mm optical path length. The oligonucleotides were diluted to 20 μM in a buffer containing 20 mM Tris pH 8 and appropriate concentration of K^+/Na^+ ions. Before each measurement, the oligonucleotides were heated to 90°C for ∼5 min and slowly cooled to room temperature, to avoid formation of intermolecular structures. An average of three scans was recorded between 220 and 320 nm at room temperature. The spectra were corrected for baseline and signal contributions from the buffer.

Vectorial GQ folding using a superhelicase Rep-X

In order to eliminate end-effects of the sequence on Rep-X unwinding and GQ folding, we introduced additional 10 nt at the 3′ end of hTel28: 5′-TGGCGACGGCAGCGAGGC**GGG(TTAGGG)$_4$T**/Cy3/GGACGCACGG-3′. The hTel28 was then annealed to its complementary sequence in the ratio of 1:1.1 at 95°C for 5 mins, in a buffer containing 50 mM NaCl and 10 mM Tris–HCl, pH 8, followed by slow cooling to room temperature. The duplexes were then immobilized on the PEG-passivated surface. A total of 50 nM Rep-X was then incubated in a loading buffer (10 mM Tris–HCl, pH 8, 10% glycerol, 1% BSA) for 2 min. The unbound Rep-X was washed off simultaneously and the unwinding reaction was initiated by adding the unwinding buffer (10 mM Tris–HCl, pH 8, 100 mM KCl, 2.5 mM $MgCl_2$, 1 mM adenosinetriphosphate (ATP) [ThermoFisher Scientific], 10% glycerol, 1% BSA). For characterization of the maiden folded GQ conformation, the unwinding reaction was quenched after a minute and images were acquired for

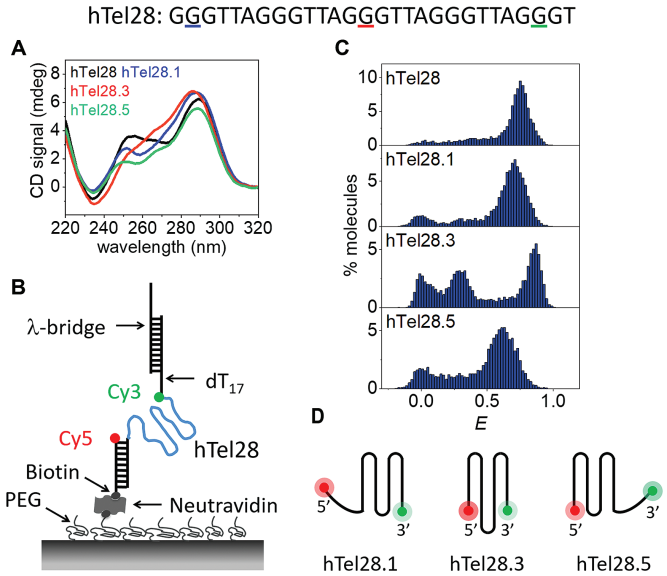

Figure 1. Conformational analysis of a five-repeat telomeric DNA (hTel28). (**A**) CD spectra of hTel28, hTel28.1, hTel28.3 and hTel28.5 in 100 mM K^+. The G to T mutation sites in hTel28 corresponding to hTel28.1, hTel28.3 and hTel28.5 are underlined in colors matching the figure legend. (**B**) Schematic diagram of hTel28 construct for smFRET. The 5′ extension to the 28 nt long human telomeric repeat $GGG(TTAGGG)_4T$ was annealed to a 18 nt long biotinylated strand and immobilized on a PEG-passivated quartz surface through biotin-neutravidin interaction. The 5′ end of the biotinylated strand is labeled with Cy5 (acceptor). FRET was measured between Cy3 and Cy5. The main hTel28 strand is labeled with Cy3 (donor) at the 3′ end of the telomeric repeat and is followed by dT_{17} and an 18 nt extension that is annealed to 30 nt long λ-bridge. (**C**) *E* histograms of hTel28, hTel28.1, hTel28.3 and hTel28.5 in 100 mM K^+ concentration. (**D**) Schematic representation of GQ formation after site-specific G to T mutation of hTel28.

smFRET analysis. For the real-time observations, imaging was started a few seconds before addition of the unwinding buffer. A detailed protocol has been described by Hua *et al.* (34).

RESULTS

smFRET analysis of five repeats human telomeric DNA

Our hTel28 construct consists of a central 28 nt long human telomeric DNA sequence spanning five hexanucleotide repeats, $(GGG(TTAGGG)_4T)$, flanked by two duplex handles at the 5′ and 3′ ends. CD spectrum of hTel28 in 100 mM K^+ showed a positive peak at ∼289 nm with shoulders at 265 and 255 nm and a negative peak at 235 nm, suggestive of hybrid GQs (Figure 1A and Supplementary Figure S1a) (41). The construct was immobilized on to the PEG-passivated surface via the duplex stem at the 5′ end whereas a 3′ end partial duplex served as a tether to the λ-DNA for use in fluorescence-force spectroscopy. The donor (Cy3) and acceptor (Cy5) fluorophores were placed adjacent to the ends of the telomeric sequence such that FRET efficiency, *E*, reflects the end-to-end distance of the telomeric sequence (Figure 1B).

Formation of secondary structures was first examined through smFRET as a function of K^+ concentration (Supplementary Figure S1b). The peak centered at $E = 0$ represents molecules with a missing or inactive acceptor, and hence can be ignored. In the absence of K^+, we observed a peak centered at $E \sim 0.16$. Transition into secondary structures commenced between 2 and 10 mM K^+ as suggested by emergence of broad peaks at $E \sim 0.3$ and ~ 0.69 at 10 mM K^+, which then resulted in a major folded population at $E \sim 0.75$ at 100 mM K^+ (Figure 1C and Supplementary Figure S1b). The populations of low and high *E*-values were attributed to unfolded and folded conformations, respectively. Single molecule trajectories showed no *E* transitions during our observation time up to ∼90 s at a time resolution of 30 ms (Supplementary Figure S1d).

In a five-repeat telomeric sequence, GQs can potentially form at the 3′ or 5′ ends by association of four consecutive T_2AG_3 repeats (3′ and 5′ GQs, respectively), or with an inner long loop bearing the second, third or fourth T_2AG_3 segment (long-loop GQs). Therefore, at least five different GQ structures can be conceived depending on which repeat is excluded (Supplementary Figure S2). In order to shed light on the conformational landscape, mutations can be introduced in biomolecules to selectively depopulate certain species from a heterogeneous population (24,29,42). Previous studies have also shown that guanine to thymine substitution of the central G of a hexanucleotide repeat T_2AG_3 precludes its participation in GQ formation in 100 mM K^+ or Na^+ (24,30). We replaced guanine of one T_2AG_3 segment in hTel28 by thymine to selectively reduce the possibility of that segment becoming part of a GQ. Five such variants of hTel28 were designated as hTel28.1, hTel28.2, hTel28.3, hTel28.4 and hTel28.5 based on the position of the segment bearing G to T mutation, counting from the 5′ end. For example, hTel28.1 would fold into a 3′ GQ because the mutated 5′ most repeat would not participate in a GQ under physiological ionic conditions.

CD spectra of the hTel28 variants in 100 mM K^+ showed signatures of hybrid GQs as was the case for hTel28 itself (Figure 1A) (41). However, they cannot individually fully recapitulate that of hTel28, suggesting co-existence of multiple species in hTel28. Folding of the hTel28 variants into secondary structures was modulated by cationic concentrations (data not shown) and in 100 mM K^+, a major high *E* population with peaks centered at ∼0.71, 0.85, 0.85, 0.86 and 0.62 was observed for hTel28.1, hTel28.2, hTel28.3, hTel28.4 and hTel28.5, respectively, (Figure 1C and Supplementary Figure S1d). Previous smFRET studies with a 22 nt GQ-forming telomeric sequence (hTel22) harboring four repeats reported a major *E* peak centered at ∼0.88 in 100 mM K^+ (30,32), similar to the peaks at $E \sim 0.85$ observed for hTel28.2, hTel28.3 and hTel28.4. Therefore, we tentatively assign $E \sim 0.85$ subpopulation of hTel28 to long-loop GQs where a telomeric repeat at the second, third or fourth position becomes a part of a loop (Supplementary Figure S2). On the other hand, GQs formed from four consecutive repeats mimicked by hTel28.1 and hTel28.5 would leave out a 6 nt single-stranded portion, adding to the separation between the fluorophores (Figure 1C), consistent with the lower *E*-values observed for hTel28.1 and hTel28.5 compared to the long-loop GQs. The hTel28 variants were very stable as evidenced by steady *E*-values (Supplementary Figure S1d). hTel28.1 and hTel28.5 differ only in that the G-rich pendant is at the 3′ and 5′ end of the sequence, re-

spectively (Figure 1D). However, the corresponding FRET distributions are significantly different (Figure 1C). As the folding pattern of GQ bearing an unstructured poly(dT) stretch is independent of whether the latter is at the 3′ or 5′ end (43), the difference in *E* between hTel28.1 and hTel28.5 likely results from different GQ conformations dependent on the relative position of the unpaired T_2AG_3 tail.

Among the variants, hTel28.1 appears closest to hTel28 in the major smFRET population's peak position, suggesting a preferential formation of GQ at the 3′ end of a five-repeat telomeric sequence (Figure 1C). However, given the significant overlap with the *E* distributions of the other four hTel28 variants and hTel28, 5′ end GQs and long-loop GQ may exist as minor populations (Figure 1C and Supplementary Figure S1d). In fact, using the peak centers and widths from *E* histograms of hTel28 variants, we were able to resolve three non-interconvertible subpopulations in hTel28 with *E*-values of ∼0.66, 0.75 and 0.85 accounting for ∼18% 5′ GQ (37 out of 200 events), 64% 3′ GQ (128 events) and 18% long-loop GQ (35 events) of the recorded events, respectively (Supplementary Figure S1c).

Vectorial folding of hTel28, mimicking telomerase action

Our smFRET studies under refolding conditions suggest predominant GQ formation at the 3′ end of hTel28s. However, because a telomerase adds one repeat at a time from the 5′ to 3′ direction to a telomeric DNA under a physiological situation of telomere elongation, a GQ may first form on the 5′ side during telomere elongation. To address this point, we employed a highly processive superhelicase Rep-X to mimic the vectorial nature of telomerase-catalyzed DNA synthesis (33–34,44). In a previous study, we used this 5′ to 3′ DNA helicase to unwind RNA/DNA heteroduplex and to reveal the RNA strand in the direction of transcription (5′ to 3′) at the speed of transcription to mimic co-transcriptional RNA folding (34,35). Here, we annealed hTel28 to its complementary C-rich DNA and Rep-X was loaded through the 3′ overhang in the absence of $MgCl_2$ and ATP (Figure 2A). Duplex unwinding was then initiated by addition of $MgCl_2$ and ATP. The helicase translocates in the 3′ to 5′ direction along the C-rich DNA strand, unwinding the duplex and releasing the G-rich DNA in the 5′ to 3′ direction, which is the direction of telomere extension. GQ folding in nascent telomeric DNA synthesized during telomerase extension has been directly observed by Jansson *et al.* (45). Likewise, in our assay the disengaged G-rich hTel28 can fold into GQ as it is being revealed starting from the 5′ side (Figure 2A).

In the fully annealed state, hTel28-duplex showed a FRET peak at $E \sim 0.07$ (Figure 2B). Rep-X mediated unwinding reaction was quenched after a minute by washing away ATP and the resulting smFRET distribution showed that ∼47% of molecules fold into a GQ with a major peak at $E \sim 0.73$ (Figure 2B). The rest stayed at $E \sim 0.07$ probably because these molecules did not get unwound. A control experiment performed with a slowly hydrolysable ATP analog, AMP-PNP, showed only ∼5% folded population (Supplementary Figure S3d). Under identical unwinding buffer conditions, major peaks centered at $E \sim 0.72$ and 0.62 were observed in hTel28.1 and hTel28.5, respectively (Supplementary Figure S3a). Hence, the major state observed at $E \sim 0.73$ was attributed to 3′ GQs.

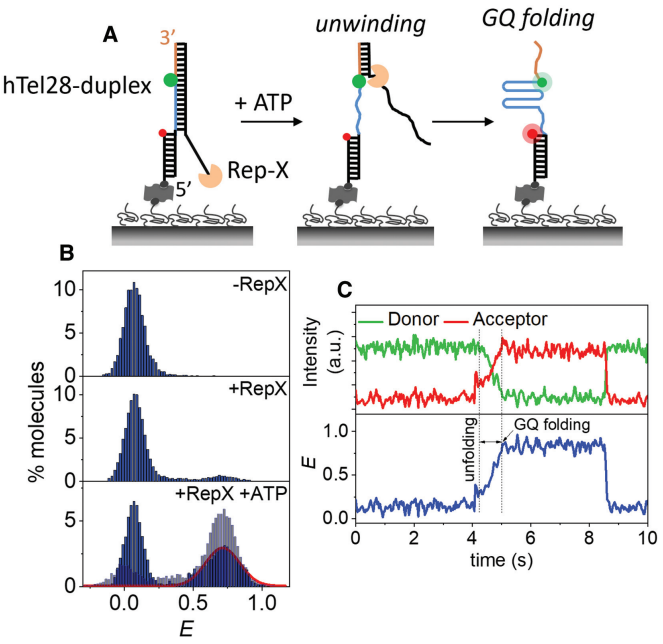

Figure 2. Vectorial folding of hTel28. (**A**) Schematic of unfolding of a hTel28-duplex via Rep-X and subsequent folding of hTel28 into GQ. (**B**) *E* histograms of hTel28-duplex in the Rep-X unbound (top) and bound (middle) states, and after Rep-X unwinding of the duplex in presence of ATP and $MgCl_2$ (bottom). The curve represents Gaussian fit to the folded GQ population. *E* histogram of hTel28.1 in the unwinding buffer is superimposed in light blue (bottom). (**C**) Representative single molecule trajectory showing unwinding of the duplex and folding of GQ (30 ms integration time).

Real-time single molecule trajectories during unwinding typically showed gradual increases in *E* to ∼ 0.73 over ∼0.7 s without pausing at intermediate FRET states (Figure 2C; Supplementary Figure S3b and c). Such behavior was observed in ∼64% (74 of 116 events) of trajectories that resulted in folded GQs. This suggests that GQ forms at the 3′ end even during vectorial folding. 5′ GQ may form transiently but if so, it is not long-lived enough to be detected as a clear intermediate. About 22% (26 events) of the folded GQs showed $E \sim 0.62$, suggestive of 5′ GQs as a minority population. Thus, although vectorial unwinding of hTel28-duplex via Rep-X exposes the T_2AG_3 segments successively from the 5′ end, GQs are formed preferentially at the 3′ end.

Dynamics of hTel28 under tension

For fluorescence-force spectroscopy, the hTel28 construct was annealed with λ-DNA via the λ-bridge prior to its immobilization on the PEG-passivated surface. The other end of the λ-DNA was tethered to an optical trapped bead (Figure 3A). The molecule under investigation was stretched and relaxed by translating the sample stage at a speed of 455 nm/s. In a typical experiment, the applied force was increased from ∼0.3 pN to ∼28 pN in ∼6.5 s, followed by relaxation at the same speed. The donor and acceptor fluorescence intensities were recorded as a function of force in each pulling cycle and this was repeated until fluorophore

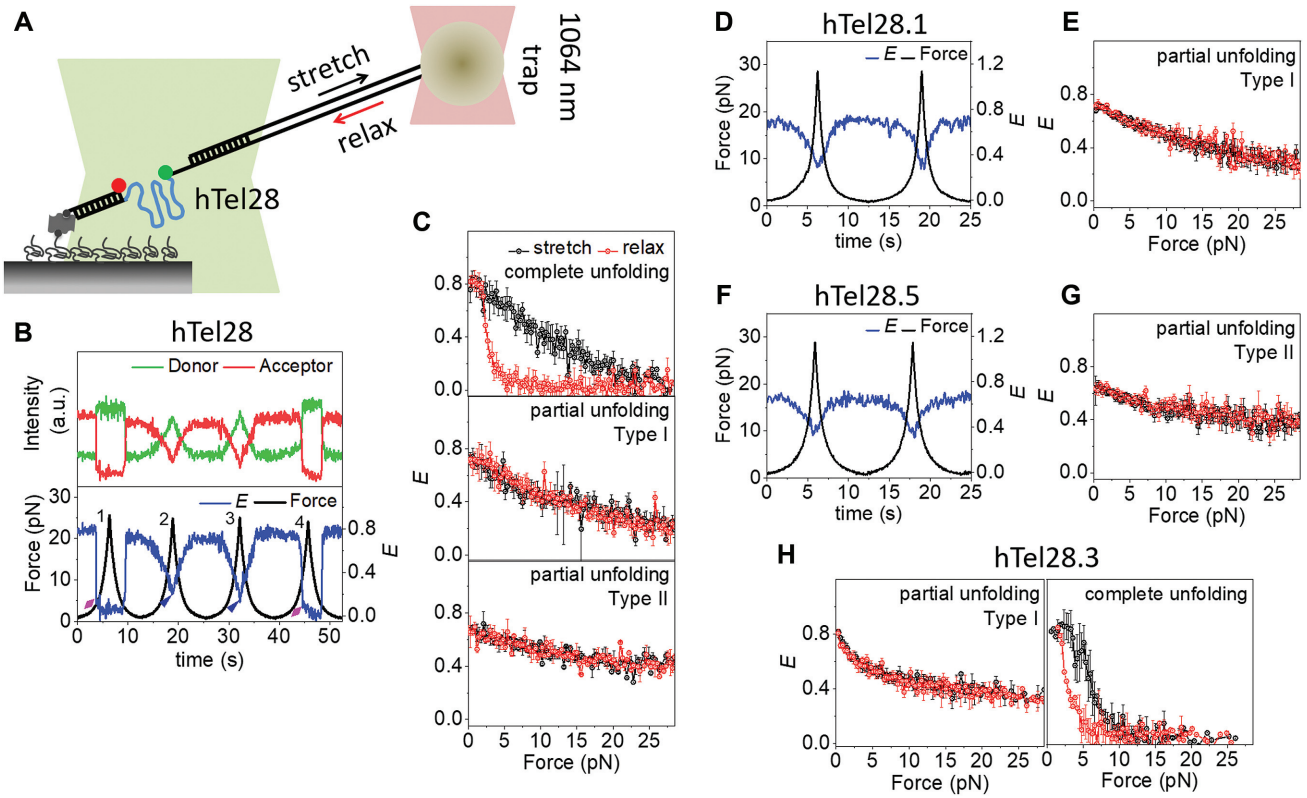

Figure 3. Conformational dynamics of five human telomeric repeats under tension in 100 mM K^+. (**A**) Schematic of integrated fluorescence-force spectroscopy assay: The hTel28 strand is annealed to a biotinylated strand and immobilized on a neutravidin coated quartz surface. The other end is connected to an optically trapped 1 μm diameter bead through a λ-DNA. Force was applied by translating the microscope stage at a speed of 455 nm/s. FRET was measured between Cy3 (donor) and Cy5 (acceptor) as a function of force. (**B**) A representative single molecule time trace of donor and acceptor intensities and corresponding *E* of hTel28. The fluorescence-force time trajectories were grouped into three classes based on the lowest *E*-value attained at a maximum applied force ∼28 pN: Complete unfolding ($E \sim 0$), Type I unfolding ($E \sim 0.2$) and Type II unfolding ($E \sim 0.37$). Complete unfolding was observed in pulling cycles 1 and 4 (diamond-arrowheads) while cycles 2 and 3 showed Type-I unfolding (triangle-arrowheads). (**C**) Average *E* versus force response from molecules undergoing complete ($N = 58$, top) partial (I) ($N = 150$, middle) and partial (II) ($N = 55$, bottom) unfolding under tension. (**D** and **F**) Representative *E* time traces of hTel28.1 (D) and hTel28.5 (F) during two pulling cycles. (E, **G** and **H**) Average *E* versus force response of hTel28.1 (E), hTel28.5 (G) and hTel28.3 (H). All single molecule trajectories were collected at a time resolution of 20 ms. Error bars represent standard errors.

photobleaching or tether rupture. Data were collected from at least 20 different molecules under each condition.

Diverse conformational changes of hTel28 in 100 mM K^+ were observed under tension. The extent of hTel28 unraveling, as judged by the lowest *E*-values achieved, varied across different stretching cycles of a single molecule (Figure 3B) and could be broadly categorized as complete unfolding (22%, 58 out of 263 events), Type I partial unfolding (∼57%, 150 events) and Type II partial unfolding (∼21%, 55 events). Typically, complete unfolding of a molecule was characterized by stable $E \sim 0.8$ up to a certain force, followed by an abrupt decrease in *E* to ~ 0.1 (Figure 3B, cycles 1 and 4, and Supplementary Figure S4a). Further gradual decreases from $E \sim 0.1$ suggests stretching of an already unfolded molecule. The subsequent relaxation event, similarly, showed an abrupt recovery to the initial *E*-value. The unfolding (f_{unfold}) and refolding forces (f_{refold}) were defined as the force(s) corresponding to midpoint(s) of the *E* transitions during abrupt unfolding and refolding. Diverse f_{unfold} values observed across different pulling cycles could be fit to two force clusters centered at ∼7.5 and 17 pN (Supplementary Figure S4b and c). Refolding, on the other hand, predominantly occurred at lower forces, between 1 and 7 pN (Supplementary Figure S4d).

The most common was Type I partial unfolding, defined by stable *E*-values of ∼0.75 up to ∼2 pN, followed by a gradual decrease to ∼0.25 at ∼28 pN (Figure 3B, cycles 2 and 3). On relaxation, the unraveling pathway was retraced without any significant hysteresis (Figure 3C). Type II partial unfolding also showed reversible *E* changes but with a different range, from ∼0.68 to ∼0.37 (Figure 3C).

Interestingly, a single molecule of hTel28 under tension can interchange between complete and Type I partial unfolding. For example, Figure 3B shows complete and partial unfolding in consecutive pulling cycles 1 and 2 and vice versa in the cycles 3 and 4. Overall, switching from complete to Type I partial unfolding in consecutive cycles and vice versa were observed in ∼33% (19 out of 58 events) and ∼16% (24 out of 150 events) pulling cycles, respectively. However, Type II partial unfolding events rarely switched to other types of unfolding in the subsequent cycle (4 events, ∼7%).

Conformational dynamics of hTel28 variants under tension

We next examined the hTel28 variants under tension in order to gain mechanistic insight into GQ formation in five-repeat telomeric DNA. In 100 mM K^+, hTel28.1 showed a stable E of ∼0.72 up to applied forces of ∼2 pN, followed by a gradual decrease in E to ∼0.25 at a force of ∼8 pN (Figure 3D and E; Supplementary Figure S5c). It retraced the same E versus force during relaxation (Figure 3E) and subsequent pulling cycles yielded similar stretching-relaxation behavior. hTel28.5 also showed reversible changes in E upon stretching and relaxation. However, the E-values changed between 0.35 and 0.62, which is narrower in range than hTel28.1 which ranged between 0.25 and 0.72 (Figure 3F and G; Supplementary Figure S5c). Conformational dynamics of hTel28.1 and hTel28.5 under tension therefore are similar to Type I and Type II partial unfolding observed in hTel28, respectively. Therefore, we propose that 3′ GQs, mimicked by hTel28.1, show Type I partial unfolding and 5′ GQs, mimicked by hTel28.5, show Type II partial unfolding.

We next studied the conformational dynamics of hTel28.3 as a representative of long-loop GQs. Because the hTel28.3 harbors a G to T mutation in the innermost T_2AG_3 segment, a GQ would form preferentially with the four outer repeats with an inner 9 nt long loop (Supplementary Figure S5a). Complete and abrupt unfolding characterized by E drop from ∼0.83 to ∼0.15 and subsequent single-step refolding at lower forces accounted for ∼48% (50 out of 104 events) of the pulling cycles (Figure 3H and Supplementary Figure S5d–g). The f_{unfold} values clustered around an average value of ∼6 pN, close to the ∼7.5 pN f_{unfold} cluster observed in complete unfolding of hTel28 (Supplementary Figure S5f and g). The remaining 52% (54 events) showed gradual, reversible changes in E from ∼0.83 to ∼0.4, similar to Type I partial unfolding of hTel28 (Figure 3H).

In all, the mechanical behavior of hTel28 was recapitulated by the hTel28 variants studied. We can attribute complete unfolding to long-loop GQs, and attribute Type II partial unfolding to 5′ GQs. Type I partial unfolding was observed in both long-loop GQs and 3′ GQs. Interestingly, among the variants, mechanical heterogeneity could be noted only in long-loop GQs (hTel28.3). Combining these results with zero-force smFRET studies, we propose that in 100 mM K^+, the 3′ GQs constitute the preferred conformation of hTel28.

Effect of an unassociated G-rich segment at the 5′ end of GQ

We previously classified diverse mechanical behavior of hTel22 that contain four telomeric repeats into (i) abrupt and complete unfolding, (ii) partial unfolding and (iii) an ultrastable population that could not be perturbed by forces up to ∼28 pN (30). GQs formed at the 3′ or 5′ ends of hTel28 would have a 6 nt overhang to an hTel22-like GQ core. Therefore, an ultrastable GQ with a pendant 6 nt tail would show gradual E decreases as the ssDNA tail is stretched. Thus, the reversible E changes between ∼0.68 and ∼0.39 under applied forces shown by 5′ GQs and classified as Type II partial unfolding of hTel28 may be due to an ultrastable GQ population. On the other hand, a greater degree of unraveling observed for Type I partial unfolding (Supplementary Figure S5c) can be due to partial unfolding via mutual slippage of the G-rich strands as proposed to occur in hTel22 (30). Interestingly, Type I partial unfolding, although dominant in hTel28, was observed only in ∼11% of hTel22 molecules (30). Hence, in order to investigate the effect of a short addendum at the 5′ end of GQ, we substituted the 6 nt T_2AG_3 segment at the 5′ end with $(dT)_6$ (hTel28.1T). Thus, hTel28.1T serves as a 22 nt long GQ forming sequence with a pendant $(dT)_6$ at its 5′ end.

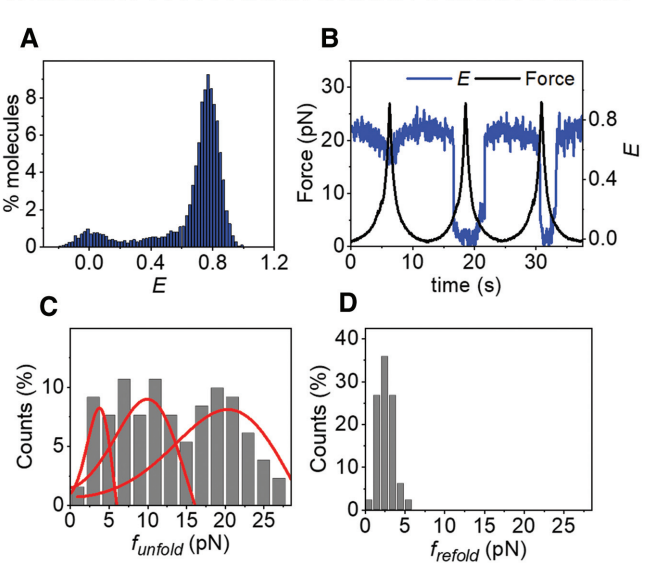

Figure 4. (**A**) E histogram of hTel28.1T in 100 mM K^+, in the absence of force. (**B**) A representative E time trace of hTel28.1T during three pulling cycles. (**C** and **D**) Distributions of f_{unfold} (C) and f_{refold} (D). Only those cycles showing complete unfolding were included ($N = 131$). The curves represent the rupture force distributions predicted using the Dudko–Szabo model (39,40).

hTel28.1T folded in a K^+ concentration-dependent manner (Supplementary Figure S6a), showing a major population at E ∼ 0.77 (Figure 4A and Supplementary Figure S6a). The E-values are similar between hTel28.1T, hTel28 and hTel28.1, further suggesting that a GQ forms predominantly at the 3′ end of a five-repeat telomeric sequence. Under tension in 100 mM K^+, hTel28.1T showed three qualitatively different types of behavior. For example, a molecule can gradually stretch to E ∼ 0.47 from an initial E ∼ 0.77 and back, likely as a result of stretching of the 6 nt ssDNA next to an ultrastable GQ (Figure 4B (cycle 1) and Supplementary Figure S6c) (46). The range of E-values is comparable with that observed in Type II partial unfolding, supporting our assignment of hTel28.5 to an ultrastable GQ. In the second and third pulling cycles shown in Figure 4B, we observed a gradual decrease in E from ∼0.77 due to ssDNA stretching, followed by abrupt unfolding (Figure 4B). Refolding also occurred abruptly (Figure 4B and D). This second type of behavior showed a wide range of unfolding forces (∼1 and 28 pN) (Figure 4C and Supplementary Figure S6b), similar to cooperative unfolding observed for hTel22 (30). The f_{unfold} values (131 events) can be grouped into three force clusters centered at ∼3.5, 10 and 20 pN (Figure 4C). We also observed a third type of unraveling behav-

ior, characterized by gradual and reversible changes from $E \sim 0.77$ to ~ 0.25 (24 events, ~10%) and reminiscent of Type I partial unfolding of hTel28 (Supplementary Figure S6d) (30). In all, we could recapitulate conformational dynamics of hTel22 with hTel28.1T. The stark contrast in mechanical behavior of hTel28.1 and hTel28.1T suggests that GQ conformations are affected by the unassociated G-rich strand at the 5′ end in a way that a $(dT)_6$ cannot mimic.

Formation of GQs in six-repeat human telomeric DNA

We next investigated the formation of GQs in six-repeat human telomeric DNA, hTel34 (Figure 5A). Like hTel28, CD spectra of hTel34 in 100 mM K^+ shows signatures of hybrid GQs (Figure 5B) (41). Secondary structure formation commenced at < 10 mM K^+ (Supplementary Figure S7a), and in 100 mM K^+, we observed two non-interconvertible (within our observation time) major populations, centered at $E \sim 0.54$ and 0.71, that we designate as E_1 and E_2 states (Figure 5C and Supplementary Figure S7b).

Under tension, E_1, which is the dominant population, unraveled gradually and reversibly (89 out of 137 events, ~65%, Figure 5D and E) whereas E_2 unfolding occurred abruptly (48 events, Figure 5D). f_{unfold} for E_2 showed a wide distribution (1–20 pN) that could be fitted to two clusters centered at ~7.5 and 16 pN (Supplementary Figure S7c and d) and refolding occurred at lower forces (Figure 5E and Supplementary Figure S7c). Switching from E_2 to E_1 after a pulling cycle accounted for ~29% (14 out of 48 events, Figure 5D (cycle 4)) of the events originating at E_2 whereas a E_1 to E_2 switch was observed in ~13% (12 out of 89 events, Figure 5D (cycle 2)) of the events initiating from E_1. A minor population at $E \sim 0.33$ (E_0) is similar in mechanical response to $(dT)_{34}$, an unstructured ssDNA of the same length (Figure 5F) and therefore is likely an unfolded population.

With two additional T_2AG_3 repeats, hTel34 can form 15 GQ structures that differ in which four of the six repeats are chosen (Supplementary Figure S2). In order to gain information on its internal architecture we placed the donor fluorophore at the thymine in the 22nd position (hTel34′, Figure 5A) so that FRET reports on the conformation of the first four repeats, counting from the 5′ end. We have previously shown that fluorescent labels do not significantly alter the folded states of GQs (30). At 100 mM K^+, the smFRET distribution showed three major non-interconvertible (within our observation time) populations centered at $E \sim 0.47$ (E_0'), 0.68 (E_1') and 0.86 (E_2') (Figure 5G and Supplementary Figure S8a and b). E_0' showed ssDNA-like mechanical response (Supplementary Figure S8c) (30,46) similar to the mid-FRET population of hTel22 attributed to an unstructured ssDNA (Supplementary Figure S8c) (30). Like E_1 of hTel34, E_1' unraveled gradually under tension (128 out of 186 events, ~68%, Figure 5H and i (top)). Like E_2 of hTel34, E_2' unfolded abruptly (58 events, Figure 5h (bottom)) at around two force clusters of ~9 and 14 pN (Supplementary Figure S8e) while refolding occurred at lower forces (Figure 5H; Supplementary Figure S8d and e). The gradual and reversible mechanical responses of the dominant populations of hTel34 (E_1) or hTel34′ (E_1') can be likened with the non-cooperative unraveling via mutual strand slippage, observed with 3′ GQs in hTel28. This further hints at preferential GQ formation at the 3′ end in hTel34, similar to hTel28.

DISCUSSION

Despite its simple repetitive sequence, telomeric GQ possesses extremely heterogeneous populations. Structural studies have revealed at least six different GQ conformations (e.g. parallel, antiparallel, hybrid, etc.) based on strand symmetry, strand orientation and glycosidic conformation (18–24). Recently, we identified six mechanically different, interconvertible GQ species in a 22 nt long human telomeric DNA using single molecule fluorescence-force spectroscopy (30). In the context of longer telomeric sequences harboring more than four T_2AG_3 repeats, as we studied here using five repeats and six repeats using fluorescence-force spectroscopy, GQ polymorphism in tandem with their positional multiplicity along the DNA add to the complexity. Indeed, previous studies of four to seven repeats using smFRET only (hence at zero force) (32) or optical tweezers only (25) found evidence of multiple states. Using a series of mutants designed to mimic positionally defined GQs, we were able to separate positional multiplicity and GQ polymorphism.

Five telomeric repeats

Here, we found that the mechanical response of GQ in a long telomeric DNA can depend on its location (Figure 6). For example, although both the 3′ GQ and the 5′ GQ in hTel28, mimicked by hTel28.1 and hTel28.5, respectively, utilize four consecutive T_2AG_3 segments, on application of force, the 3′ GQ former unravels gradually whereas the 5′ GQ withstands forces as high as 28 pN (Supplementary Figure S5c). This may stem from differences in the underlying GQ conformations as evidenced by their zero force E-values (~0.72 for hTel28.1 and ~0.62 for hTel28.5, Supplementary Figure S5b). Because a GQ bearing a poly(dT) tail at the 5′ end had identical FRET value to a GQ with the same poly(dT) tail at the 3′ end (47), the different E-values between hTel28.1 and hTel28.5 likely come from structurally different GQs (Supplementary Figure S5a and b). The gradual unfolding observed for 3′ GQ suggests a compliant structure, held together by local interactions. Under tension, due to relative motion of the bases and backbone, the terminal base pairs are frayed leading to vertical slippage of a G-rich strand away from the native stem. Thus, individual G-quartets are disrupted one at a time, and there is a stepwise reduction in the number of G-quartets during unfolding (48,49). Such a mode of unraveling via strand slippage has been described for parallel GQs as slippage of a G-strand does not perturb the inherent all *anti*-G configuration of the remaining G-quartets. However, as antiparallel or hybrid GQs are each characterized by a unique arrangement of *anti* and *syn*-Gs, strand slippage upsets the inherent G arrangement in a quartet and disarrays the entire structure cooperatively.

Among the variants of hTel28, only a long-loop variant, hTel28.3, showed cooperative destabilization of the G-quartets, marked by abrupt unfolding under increasing force. Analysis of the corresponding f_{unfold} histogram, which

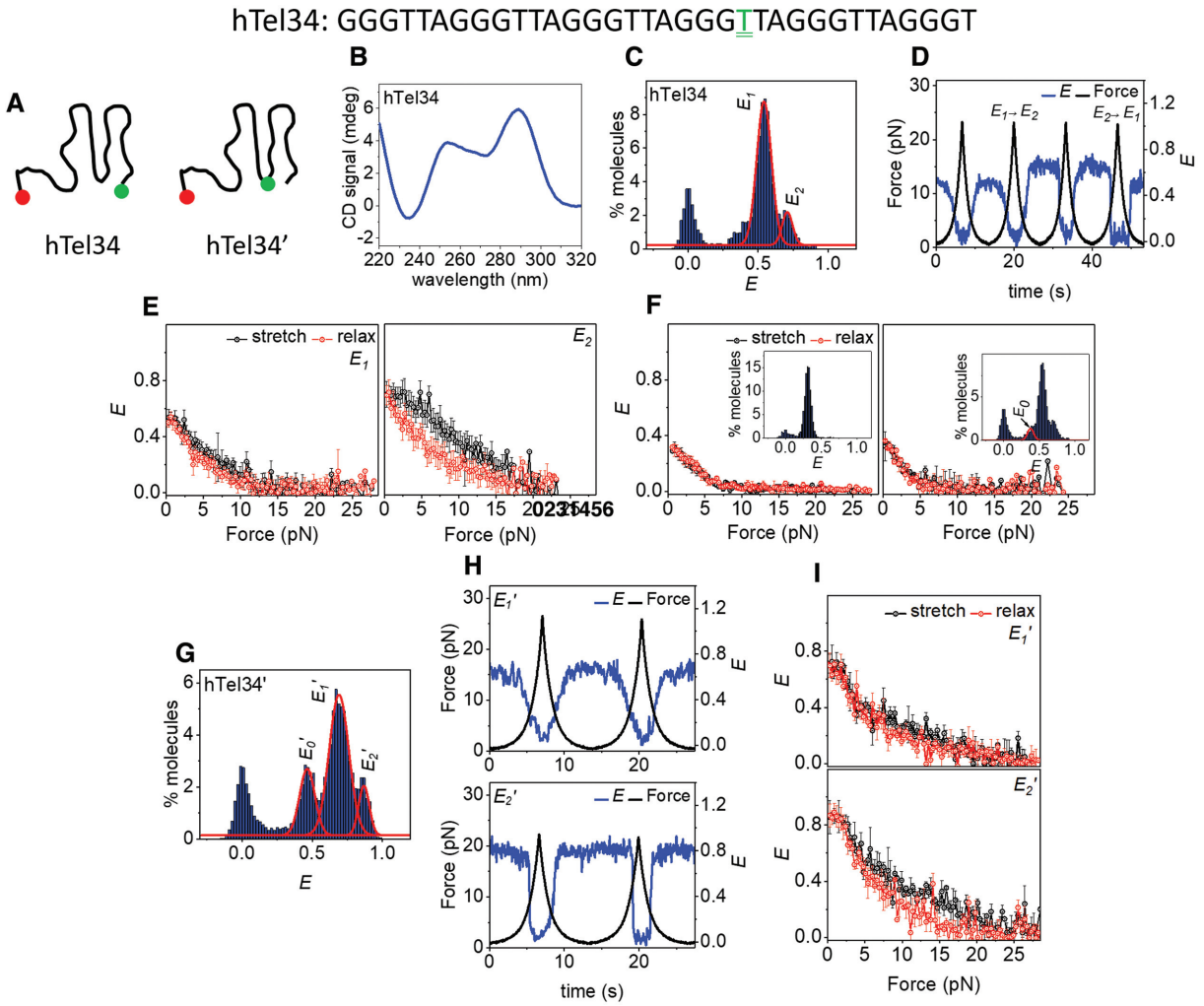

Figure 5. Conformational analysis of hTel34 in 100 mM K^+. (**A**) Schematic representations of hTel34 and htel34′. T_{22} bearing the donor dye in hTel34′ is doubly underlined in green. (**B**) CD spectra of hTel34 in 100 mM K^+. (**C**) *E* histogram of hTel34 in the absence of force. The curves denote E_1 and E_2 populations. (**D**) A representative *E* time trace of hTel34 during four pulling cycles. (**E**) Average *E* versus force plots of E_1 ($N = 89$, left) and E_2 ($N = 48$, right). (**F**) Average *E* versus force response of $(dT)_{34}$ (top) and E_0 of hTel34 (bottom). The corresponding *E* histograms in the absence of force are shown in inset. (**G**) *E* histogram of hTel34′ in the absence of force. The curves denote the E_0', E_1' and E_2' populations. (**H**) Representative *E* time traces of E_1' (top) and E_2' (bottom) of hTel34′ during two pulling cycles. (**I**) Average *E* versus force plots of E_1' ($N = 128$, top) and E_2' ($N = 58$, bottom). Error bars represent standard errors.

is peaked at 7.5 pN, yielded free energy parameters $\Delta x^{\ddagger}$ (distance to the transition state) and $\Delta G^{\ddagger}$s (apparent free energy of activation) as ~6.6 nm and ~7 k_BT, respectively (Supplementary Figure S9a) (39,40). In contrast, complete unfolding of hTel28 showed two f_{unfold} clusters at ~7.5 and 17 pN, corresponding to $\Delta x^{\ddagger}$s of ~6.7 and 3.2 nm, respectively (Supplementary Figure S9a). A $\Delta x^{\ddagger}$ of ~6.1 nm was estimated for a mechanically weak subpopulation (f_{unfold} < 5 pN) of hTel22 under identical buffer conditions (30), attributed to partially folded GQs containing two G-quartets instead of three (30). Because $\Delta x^{\ddagger}$ of the 7.5 pN cluster of hTel28 and hTel28.3 is close to that of the two-tiered hTel22-GQs, we suggest the presence of frayed G-quartet structures adjacent to the inner T_2AG_3-containing loop. The 17 pN cluster can be attributed to hTel28.2 or hTel28.4 because it was not seen for hTel28.1, hTel28.3 or hTel28.5.

At applied forces of up to ~28 pN, because the 3′ GQs in hTel28 are only partially unraveled, they rarely convert into long-loop GQs (~16%) upon relaxation. In contrast, a long-loop GQ, when unfolded cooperatively, loses memory of its initial state and can relax into a mechanically different long-loop GQ or even a 3′ GQ (~33%).

Although our single molecule studies suggest preferential GQ formation at the 3′ end of hTel28, such GQs with a 5′ T_2AG_3 overhang were mechanically distinct from GQs formed in hTel22 (Figure 6). Partial and gradual unfolding attributed to mutual slippage, which was a minority behavior in hTel22, dominates the overall mechanical response of hTel28 (30). Interestingly, upon substitution of the hexanucleotide G-rich strand with a poly(dT) segment of the same length (hTel28.1T), the canonical hTel22 like behavior was restored. In addition to ultrastable (~35%) and par-

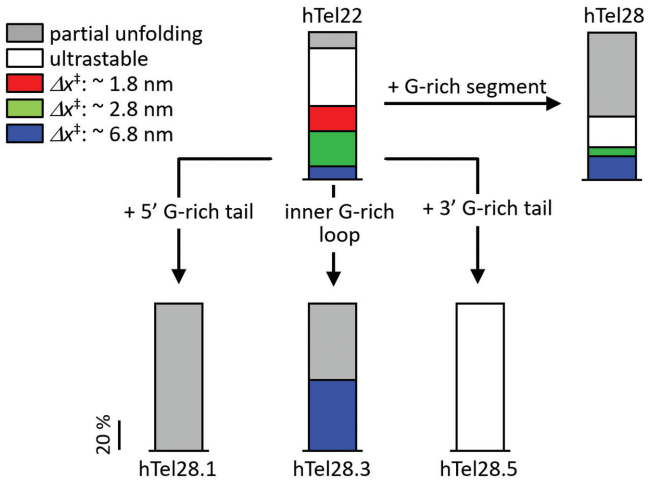

Figure 6. Effect of an unpaired G-rich segment in canonical GQs (hTel22) in 100 mM K^+. Mechanical response of hTel28, which is hTel22 with a G-rich segment, is a cumulative effect of hTel28.1, hTel28.5, hTel28.3 etc.

tially unfolded GQs (~10%), we observed at least three cooperatively unfolding species with f_{unfold} s of ~3.5, 10 and 20 pN from hTel28.1T. The respective $\Delta x^{\ddagger}$s were estimated as ~6.8, 2.8 and 1.8 nm (Figure 6 and Supplementary Figure S9a), close to the previously reported values of ~6.1, 3.1 and 1.8 nm in hTel22 (30).

Streamlining effects of an unassociated telomeric repeat on GQ folding

In all, our study highlights the regulatory role of an unassociated G-rich strand on the GQ core of a five-repeat telomeric sequence. A T_2AG_3 overhang at the 5′ end biases the molecule toward the partial unfolding pathway whereas when present at the 3′ end it renders GQ ultrastable. On the other hand, an inner loop harboring a T_2AG_3 segment manifests cooperative unfolding similar to canonical hTel22-GQs (30). We suggest that position-dependent transient interactions between the unpaired T_2AG_3 segment with the underlying GQ core simplify the conformational distribution, selecting a unique mechanical conformation out of diverse populations (Figure 6). As we will show below, this streamlining effect seems unique to potassium solutions because we did not observe such an effect with sodium solution. Interactions between unpaired G-nucleotides and the GQ core were also observed in the GQ-forming c-*Myc* promoter (50). However, detailed structural studies are required to understand the exact nature of such interactions. In cells, telomere-binding proteins such as POT1 (47) may sequester the excess DNA surrounding the GQ core such that mechanical heterogeneity can be restored. Future studies in this direction would be beneficial for comprehensive understanding of the interplay between telomeric DNA and associated proteins.

Six telomeric repeats

From hTel34 which contains six human telomeric repeats, we observed two major populations E_1 and E_2 at zero force. To reduce the compounding effect of stretching of ssDNA tail(s), we also designed the internally labeled construct hTel34′ whose FRET response is sensitive to the conformation of the four 5′ most repeats. By comparative analysis of responses of hTel28 and hTel34, we assigned the major population, E_1 (or E_1') of hTel34 (or hTel34′) to 3′GQ with a 5′ 12 nt ssDNA overhang harboring two T_2AG_3 segments. To test this assignment further, we next evaluated the possibility that E_1 (or E_1') instead represents an internal GQ formed from the second, third, fourth and fifth T_2AG_3 segments, leaving 6 nt overhangs at both the 3′ and 5′ ends (Supplementary Figure S2). Indeed, GQ formed from an internal register of telomeric repeats with 6 nt poly(dT)$_6$ tails at both the 5′ and 3′ ends yields a FRET distribution centered at $E \sim 0.5$, close to E_1(Supplementary Figure S7e). Therefore, the zero force FRET value alone cannot rule out an internal GQ. However, an internal GQ would undergo Type I unfolding if dominated by the 5′ extra T_2AG_3 segment. Under the hTel34′ labeling scheme, the placement of probes for an internal GQ is similar to that in hTel28 3′ GQ, i.e. a compact core with a 5′ tail (Supplementary Figure S7f, left). Thus, Type I unfolding would emulate E versus force response hTel28.1T (or hTel28), which is clearly not the case as shown in Supplementary Figure S7f ($E_{hTel34',20\,pN} \sim$ 0.07 versus $E_{hTel28,20\,pN} \sim 0.28$). Likewise, Type II unfolding would occur if an internal GQ is dominated by the 3′ extra T_2AG_3 segment but $E_{20\,pN}$ for Type II unfolding of 3′ GQ of hTel28 is ~0.43, very different from $E_{hTel34',20\,pN}$ ~0.07. Thus, the mechanical response of an internal GQ is inconsistent with that of E_1'-hTel34′. This further substantiates our assignment of E_1 of hTel34 to 3′GQs.

The minor populations E_2 (hTel34) and E_2' (hTel34′) show abrupt, cooperative unfolding under tension similar to long-loop GQs in hTel28. The f_{unfold} of E_2 reveals two force clusters at ~8 and 16 pN with underlying $\Delta x^{\ddagger}$s of ~3.7 and 4 nm, respectively (Supplementary Figure S9b). Although similar in $\Delta x^{\ddagger}$, the cluster at $f_{unfold} = 16$ pN had significantly higher zero-force dwell time, $\tau_u(0)$ and $\Delta G^{\ddagger}$ (10 990 versus 59.83 s and 11.6 versus 5.7 k_BT), indicating the presence of long and short-lived species that are structurally similar (24). The f_{unfold} clusters in E_2' have $\Delta x^{\ddagger}$ values of ~3.6 and 3.7 nm, similar to E_2. Hence, we suggest that E_2 of hTel34 are E_2' of hTel34′ represent the same conformation (Supplementary Figure S9b).

GQs formed at the 5′ end of hTel28 were ultrastable. Analogous behavior might be overshadowed by ssDNA stretching in hTel34, but should be reflected as stable high FRET over the complete pulling cycle in hTel34′. Although E_2' showed a no-force FRET value of ~0.88, expected for a 5′ GQ, it did not show stable high FRET under tension. Rather, E_2' showed abrupt unfolding, reminiscent of long-loop GQs in hTel28. So we suggest that E_2 of hTel34 is unlikely due to a 5′ GQ. Can it be due to a 3′ GQ? A 3′ GQ would leave a 12 nt ssDNA region in hTel34′ which can be stretched under tension, giving rise to initial gradual FRET decrease. Because we did not observe any FRET decrease until abrupt unfolding, we can also rule out the possibility that E_2' and E_2 correspond to a 3′ GQ. Overall, our data suggest that E_2 represents long-loop populations although we cannot tell whether there is a single long loop of 15 nt in length or two separate 9 nt long loops. The difference

in transition distances between hTel34 and hTel28 may be attributed to structural differences stemming from the size and arrangement of the intervening loop(s).

Dynamics in Na^+ solution

Instead of hybrid GQ signatures in 100 mM K^+, CD spectra of hTel28 in 100 mM Na^+ showed signatures of antiparallel GQs (Supplementary Figure S10a) (51). In the absence of force, we observed a major population at $E \sim 0.72$ (Supplementary Figure S10b). Under tension, unfolding predominantly occurred via an abrupt change in E from ∼0.72 to ∼0.1 at forces peaked at ∼8 pN (Supplementary Figure S11a, b and d) and refolding occurred at lower forces of ∼2.5 pN (Supplementary Figure S11e). We estimated $\Delta x^{\ddagger}$, $\tau_u(0)$ and $\Delta G^{\ddagger}$ as ∼4.3 nm, ∼50 s and 6.5 k_BT, respectively. Additionally, Type II partial unfolding was observed in ∼30% (31 out of 104 events) of pulling cycles (Supplementary Figure S11a (cycle 3) and b). We did not observe Type I partial unfolding. Among the variants of hTel28, only hTel28.1 showed stable folding in Na^+, suggesting that 3′ GQ formation is favored also in Na^+ (Supplementary Figure S10b). Other variants did not form stable GQs in Na^+ likely because of poorer coordination by smaller Na^+ compared to K^+ (52).

Under tension, hTel28.1 predominantly showed Type II partial unfolding in 100 mM Na^+ (87 out of 114 events, ∼76%) (Supplementary Figure S11f and g) with the rest showing complete unfolding with two f_{unfold} clusters centered at ∼5 and 19 pN (Supplementary Figure S11f–i). Upon substituting the hexanucleotide G-rich segment at the 5′ end with $(dT)_6$, we predominantly observed two f_{unfold} clusters corresponding to long-lived ($\tau_u(0) \sim 1190$ s, $f_{unfold} \sim 14.8$ pN) and short-lived ($\tau_u(0) \sim 68$ s, $f_{unfold} \sim 6.8$ pN) isostructural species with $\Delta x^{\ddagger}$ of ∼4 nm, which is close to a previously reported value for hTel22 (Supplementary Figure S12) (30). The rest (∼20%, 22 out of 110 events) showed Type II partial unfolding. Overall, hTel28.1T was very similar to hTel28, both in its zero-force E-value and its mechanical response. Thus, the GQ-core in Na^+ remains unaffected by an additional T_2AG_3 overhang at the 3′ end.

hTel34 adopted anti-parallel GQ-like conformation in 100 mM Na^+ and showed a single population centered at $E \sim 0.5$ (Supplementary Figure S13a and b) (51). The mechanical response of hTel34 was gradual and reversible similar to E_1 in 100 mM K^+ (Supplementary Figure S13c and d). hTel34′ manifested a major FRET population at $E \sim 0.72$ (Supplementary Figure S12e) and showed abrupt unfolding with three f_{unfold} clusters of ∼4, 9 and 19 pN (Supplementary Figure S13h) and refolding at lower forces (Supplementary Figure S13f, g and i). Prior to the abrupt decreases in E we observed gradual decrease in E with increasing force, likely due to the presence of ssDNA region within folded hTel34′. We estimated similar $\Delta x^{\ddagger}$s in 3′ GQ forming hTel28 and hTel34′ (Supplementary Figure S14). We note that a 5′ GQ in hTel34′ would resemble hTel22 in FRET values with a main population at $E \sim 0.85$ (30). Because hTel34′ shows a major conformation with $E \sim 0.72$, we can eliminate the possibility of preferential GQ formation at the 5′ end in hTel34 (or hTel34′). A long-loop GQ in hTel34′, if it exists, would harbor a 15 nt long loop or two 9 nt long loops and have $E \sim 0.85$ of hTel22 instead of the observed $E \sim 0.72$ (30). Moreover, because long-loop GQs could not be stably formed in hTel28 in 100 mM Na^+, we can rule out possible long-loop GQ formation in hTel34 and hTel34′. Thus, similar to 100 mM K^+, our data are most consistent with GQ formation predominantly at the 3′ end of hTel34 in 100 mM Na^+.

Taken together, our study suggests a complex, sequence dependent interplay between GQ and the unpaired Gs in five and six-repeat telomeric DNA. We note that, the predominant Type I partial unfolding observed in hTel28 and hTel34 is exclusive to 100 mM K^+. Thus, although GQ formation is polarized at the 3′ end in Na^+ and K^+, the regulatory effect of an associated telomeric repeat on GQ folding appears unique to K^+ conditions.

Concluding remarks

G-rich telomeric sequences are well known to form stable GQs *in vitro*. In cells, telomeric DNA spans over 100–200 nt, beyond the double-stranded region, in the form of multiple G-rich hexanucleotide repeats. Using single molecule fluorescence-force spectroscopy, we demonstrated preferential GQ formation at the 3′ end of telomeres under physiologically relevant conditions. Although the sequences used in this study can harbor only one GQ, our results might imply a general scheme of polarized GQ formation being initiated from the 3′ end of long telomeric DNA, spanning multiple repeats of GQ-forming segments. The relative position of GQ(s) with respect to the 3′ end, i.e. the size of 3′ ssDNA tail downstream of the 3′ most GQ, has important implications in processes such as extension by telomerase and alternate lengthening of telomeres (ALT). Wang *et al.* showed that a minimum 3′ ssDNA tail of ∼6, 8 and 12 nt in length is essential for helicase unwinding of telomeric GQ, extension by telomerase, and lengthening by ALT, respectively (14,53–55). Thus, GQ formation at the 3′ end of the telomeric DNA is likely to inhibit the aforementioned processes. On the other hand, although a population minority, the long-loop GQ species might be more susceptible to helicase-catalyzed by providing an internal loading site for helicases (56). Telomeres are assumed to be protected by T-loop formation, which involves invasion into the double-stranded telomeric region by the G-rich telomeric overhang and subsequent annealing with the C-rich strand (57). Stable GQs at the 3′ end might interfere with T-loop formation as well.

In the cellular milieu, processes such as transcription and replication generate tension. Although folding properties under tension have been studied quite extensively telomeric DNA containing four T_2AG_3 repeats (30,58–60), conformational diversity of longer telomeric DNA has been difficult to reveal due to positional multiplicity. We recently showed extreme heterogeneity in a four repeat telomeric DNA (hTel22) with at least six mechanically diverse, yet interconvertible species (30). In this work, the presence of additional G-rich repeat(s) was shown to simplify the folding landscape of the GQ core into a mechanically unique conformation. We suggest that the conformational diversity of GQs may be greatly reduced in telomeric DNA, which may span over more than four T_2AG_3 repeats *in vivo*. Our

data also showed that the long telomeric repeats cannot be fully unwound at average stall forces of motor proteins and polymerases (61,62). Thus, GQs can act as potential kinetic traps and inhibit replication and transcription.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Momčilo Gavrilov for providing Rep-X.

FUNDING

National Science Foundation Grant [PHY-1430124 to T.H.]; National Institutes of Health Grants [GM122569 to T.H.]; Howard Hughes Medical Institute (to T.H.). The open access publication charge for this paper has been waived by Oxford University Press – *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.
Conflict of interest statement. None declared.

REFERENCES

1. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 2013–2018.
2. Williamson,J.R. (1993) Guanine quartets. *Curr. Opin. Struct. Biol.*, **3**, 357–362.
3. Campbell,N.H. and Neidle,S. (2012) G-Quadruplexes and metal ions. *Met. Ions Life Sci.*, **10**, 119–134.
4. Neidle,S. and Parkinson,G.N. (2003) The structure of telomeric DNA. *Curr. Opin. Struct. Biol.*, **13**, 275–283.
5. Murat,P. and Balasubramanian,S. (2014) Existence and consequences of G-quadruplex structures in DNA. *Curr. Opin. Genet. Dev.*, **2014**, 22–29.
6. Taylor,J.P. (2014) Neurodegenrative diseases: G-quadruplex poses quadruple threat. *Nature*, **507**, 175–177.
7. M´etifiot,M., Amrane,S., Litvak,S. and Andreola,M.-L. (2014) G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res.*, **42**, 12352–12366.
8. Makarov,V.L., Hirose,Y. and Langmore,J.P. (1997) Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell*, **88**, 657–666.
9. Maizels,N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
10. Phan,A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.*, **277**, 1107–1117.
11. Blackburn,E.H. (1991) Structure and function of telomeres. *Nature*, **350**, 569–573.
12. Bochman,M.L., Paeschke,K. and Zaikan,V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, **13**, 770–780.
13. Lipps,H.J. and Rhodes,D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.
14. Wang,Q., Liu,J.-Q., Chen,Z., Zheng,K.-W., Chen,C.-Y., Hao,Y.-H. and Tan,Z. (2011) G-quadruplex formation at the 3′ end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res.*, **39**, 6229–6237.
15. Lv,L., Guo,Z., Wang,J. and Wang,E. (2012) G-quadruplex as signal transducer for biorecognition event. *Curr. Pharm. Des.*, **18**, 2076−2095.
16. Collie,G.W. and Parkinson,G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicine. *Chem. Soc. Rev.*, **40**, 5867–5892.
17. Neidle,S. (2010) Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.*, **277**, 1118–1125.
18. Dai,J., Punchihewa,C., Ambrus,A., Chen,D., Jones,R.A. and Yang,D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res.*, **35**, 2440–2450.
19. Lim,K.W., Amrane,S., Bouaziz,S., Xu,W., Mu,Y., Patel,D.J., Luu,K.N. and Phan,A.T. (2009) Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.*, **131**, 4301–4309.
20. Lim,K.W., Ng,V.C., Martín-Pintado,N., Heddi,B. and Phan,A.T. (2013) Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.*, **41**, 10556–10562.
21. Luu,K.N., Phan,A.T., Kuryavyi,V., Lacroix,X. and Patel,D.J. (2006) Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.
22. Parkinson,G.N., Lee,M.P. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
23. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263–282.
24. Lee,J.Y., Okumas,B., Kim,D.S. and Ha,T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 18938–18943.
25. Koirala,D., Ghimire,C., Bohrer,C., Sannohe,Y., Sugiyama,H. and Mao,H. (2013) Long-loop G-quadruplexes are misfolded population minorities with fast transition kinetics in human telomeric sequences. *J. Am. Chem. Soc.*, **135**, 2235–2241.
26. Tang,J., Kan,Z.-Y., Yao,Y., Wang,Q., Hao,Y.-H. and Tan,Z. (2008) G-quadruplex preferentially forms at the very 3′ end of vertebrate telomeric DNA. *Nucleic Acids Res.*, **36**, 1200–1208.
27. Wang,H., Nora,G.J., Ghodke,H. and Opresko,P.L. (2011) Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. *J. Biol. Chem.*, **286**, 7479–7489.
28. Xu,Y., Ishizuka,T, Kurabayashi,K. and Komiyama,M. (2009) Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. *Angew. Chem. Int. Ed.*, **48**, 7833–7836.
29. Yue,D.J., Lim,K.W. and Phan,A.T. (2011) Formation of (3+1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats. *J. Am. Chem. Soc.*, **133**, 11462–11465.
30. Mitra,J., Makurath,M., Ngo,T.T.M., Troitskaia,A., Chemla,Y.R. and Ha,T. (2019) Extreme mechanical diversity of human telomeric DNA revealed by fluorescence-force spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8350–8359.
31. Phan,A.T. and Mergny,J.L. (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.*, **30**, 4618–4625.
32. Hwang,H., Kreig,A., Calvert,J., Lormand,J., Kwon,Y., Daley,J.M., Sung,P., Opresko,P.L. and Myong,S. (2014) Telomeric overhang length determines structural dynamics and accessibility to telomerase and ALT-associated proteins. *Structure*, **22**, 842–853.
33. Arslan,S., Khafizov,R., Thomas,C.D., Chemla,Y.R. and Ha,T. (2015) Engineering of a superhelicase through conformational control. *Science*, **348**, 344–347.
34. Hua,B., Panja,S., Wang,Y., Woodson,S.A. and Ha,T. (2018) Mimicking co-transcriptional RNA folding using a superhelicase. *J. Am. Chem. Soc.*, **140**, 10067–10070.
35. Mitra,J. and Ha,T. (2019) Nanomechanics and co-transcriptional folding of Spinach and Mango. *Nat. Commun.*, **10**, 4318–4327.
36. Roy,R., Hohng,S. and Ha,T. (2008) A practical guide to single-molecule FRET. *Nat. Methods*, **5**, 507–516.
37. Hohng,S., Zhou,R., Nahas,M.K., Yu,J., Schulten,K., Lilley,D.M.J. and Ha,T. (2007) Fluorescence-force spectrocopy maps two-dimnensional reaction landscape of the holliday junction. *Science*, **318**, 279–283.
38. Zhou,R., Schlierf,M. and Ha,T. (2010) Force-fluorescence spectroscopy at the singl- molecule level. *Methods Enzymol.*, **475**, 405–426.
39. Dudko,O.K., Hummer,G. and Szabo,A. (2006) Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, **96**, 108101–108104.

40. Dudko,O.K., Hummer,G. and Szabo,A. (2008) Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 15755–15760.
41. Ambrus,A., Chen,D., Dai,J., Bialis,T., Jones,R.A. and Yang,D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
42. Zhang,Z., Dai,J., Veliath,E., Jones,R.A. and Yang,D. (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.*, **38**, 1009–1021.
43. Budhathoki,J.B., Ray,S., Urban,V., Janscak,P., Yodh,J.G. and Balci,H. (2014) RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res.*, **42**, 11528–11545.
44. Blackburn,E.H. (2005) Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS Lett.*, **579**, 859–862.
45. Jansson,L.I., Hentschel,J., Parks,J.W., Chang,T.R., Lu,C., Baral,R., Bagshaw,C.R. and Stone,M.D. (2019) Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 9350–9359.
46. Maffeo,C., Ngo,T.T.M., Ha,T. and Aksimentiev,A. (2014) A Coarse-grained model of unstructured single-stranded DNA derived from atomistic simulation and single-molecule experiment. *J. Chem. Theory Comput.*, **10**, 2891–2896.
47. Ray,S., Bandaria,J.N., Qureshi,M.H., Yildiz,A. and Balci,H. (2014) G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2990–2995.
48. Stadlbauer,P., Krepl,M., Cheatham-III,T.E., Koča,J. and Šponer,J. (2013) Structural dynamics of possible late-stage intermediates in folding of quadruplex DNA studied by molecular simulations. *Nucleic Acids Res.*, **41**, 7128–7143.
49. Stefl,R., Cheatham,T.E. 3rd, Spacková,N., Fadrná,E., Berger,I., Koca,J. and Sponer,J. (2003) Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates. *Biophys. J.*, **85**, 1787–1804.
50. You,H., Wu,J., Shao,F. and Yan,J. (2015) Stability and kinetics of c-MYC promoter G-quadruplexes studied by single-molecule manipulation. *J. Am. Chem. Soc.*, **137**, 2424–2427.
51. Dapic,V., Abdomerovic,V., Marrington,R., Peberdy,J., Rodger,A., Trent,J.O. and Bates,P.J. (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res.*, **31**, 2097–2107.
52. Lane,A.N., Chaires,J.B., Gray,R.D. and Trent,J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.
53. Fry,M. and Loeb,L.A. (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.*, **274**, 12797–12802.
54. Mendoza,O., Bourdoncle,A., Boulé,J.-B., Brosh,R.M. Jr and Mergny,J.-L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.*, **44**, 1989–2006.
55. Sun,H., Karow,J.K., Hickson,I.D. and Maizels,N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.*, **273**, 27587–27592.
56. Mohaghegh,P., Karow,J.K., Brosh,R.M. Jr, Bohr,V.A. and Hickson,I.D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.*, **29**, 2843–2849.
57. Lange,T.D. (2004) T-loops and the origin of telomeres. *Nat. Rev. Mol. Cell. Biol.*, **5**, 323–329.
58. Dhakal,S., Cui,Y., Koirala,D., Ghimire,C., Kushwaha,S., Yu,Z., Yangyuoru,P.M. and Mao,H. (2013) Structural and mechanical properties of individual human telomeric G-quadruplexes in molecularly crowded solutions. *Nucleic Acids Res.*, **41**, 3915–3923.
59. Koirala,D., Dhakal,S., Ashbridge,B., Sannohe,Y., Rodriguez,R., Sugiyama,H., Balasubramanium,S. and Mao,H. (2011) A single molecule platform for investigation of interactions between G-quadruplexes and small molecule ligands. *Nat. Chem.*, **3**, 782–787.
60. Long,X., Parks,J.W., Bagshaw,C.R. and Stone,M.D. (2013) Mechanical unfolding of human telomere G-quadruplex DNA probed by integrated fluorescence and magnetic tweezers spectroscopy. *Nucleic Acids Res.*, **41**, 2746–2755.
61. Herbert,K.M., Greenleaf,W.J. and Block,S.M. (2008) Single-molecule studies of RNA polymerase: motoring along. *Annu. Rev. Biochem.*, **77**, 149–176.
62. Yin,H., Wang,M.D., Svoboda,K., Landick,R., Block,S.M. and Gelles,J. (1995) Transcription against an applied force. *Science*, **270**, 1653–1657.