Long repeating (TTAGGG)n single stranded DNA self-condenses into compact beaded filaments stabilized by G-quadruplex formation

Anirban Kar1,*, Nathan Jones2,3,*, N. Özlem Arat1,4,*, Richard Fishel2,3,†, and Jack D. Griffith1,†

1 Lineberger Comprehensive Cancer Center, University of North Carolina-Chapel Hill, Chapel Hill, NC 27599
2 Department of Cancer Biology and Genetics, The Ohio State University Wexner Medical Center, Columbus Ohio 43210
3 The Interdisciplinary Biophysics Graduate Program, The Ohio State University, Columbus, OH 43210
4 Institute for Research in Immunology and Cancer, University of Montreal, Montreal Canada, QC H3T1J4

Running Title: Telomeric G-Strand Repeats Form Beaded Filaments

* These authors contributed equally to this work
† Correspondence may be addressed to: Jack Griffith (jdq@med.unc.edu, 919-966-8563) or Richard Fishel (fishel.7@osu.edu, 614-292-2484)

Abstract
Conformations adopted by long stretches of single stranded DNA (ssDNA) are of central interest in understanding the architecture of replication forks, R loops, and other structures generated during DNA metabolism in vivo. This is particularly so if the ssDNA consists of short nucleotide repeats. Such studies have been hampered by the lack of defined substrates greater than ~150 nt, and the absence of high-resolution biophysical approaches. Here we describe the generation of very long ssDNA consisting of the mammalian telomeric repeat (5’-TTAGGG-3’)n as well as the interrogation of its structure by electron microscopy (EM) and single molecule magnetic tweezers (smMT). This repeat is of particular interest as it contains a run of 3 contiguous guanine residues capable of forming G quartets as ssDNA. Fluorescent-dye exclusion assays confirmed that this G-strand ssDNA forms ubiquitous G-quadruplex folds. EM revealed thick bead-like filaments that condensed the DNA ~12 fold. The bead-like structures were 5 nm and 8 nm in diameter and linked by thin filaments. The G-strand ssDNA displayed initial stability to smMT force extension that ultimately released in steps that were multiples ~28 nm at forces between 6-12 pN; well below the >20 pN required to unravel G-quadruplexes. Most smMT steps were consistent with the disruption of the beads seen by EM. Binding by RAD51 distinctively altered the force extension properties of the G-strand ssDNA, suggesting a stochastic G-quadruplex-dependent condensation model that is discussed.
INTRODUCTION

DNA replication, transcription, recombination and repair all involve the generation of single stranded DNA (ssDNA) segments (1-3). In some cases, such as resection of DNA ends during repair or creation of an uncoupled replication fork, these ssDNA segments potentially could span several hundred nucleotides or more in length (4). Mixed sequence ssDNA segments are assumed to behave as random coils with some (weak) secondary structure due to local base pairing (5,6). However, if the ssDNA is long and consists of arrays of nucleotide repeats that could fold into macroscopic structures this could profoundly affect subsequent events of DNA metabolism. One well known example is the trinucleotide repeats linked to the triplet repeat diseases in which it is thought that as the replication fork passes these repeats, the ssDNA can fold into long hairpin structures and generate repeat expansions. The telomeric hexanucleotide repeat (5’-TTAGGG-3’) found in all mammals and many eukaryotes (7) provides another example of a repeating sequence with the potential of folding into large macroscopic structures when present in a ss state. The telomeric repeat is a member of a subset of repetitive DNAs found widely across nature containing runs of 3 to 4 contiguous guanine bases that as ssDNA, have the ability to form G-quadruplexes (8). Other examples include the replication origin of Kaposi’s Sarcoma-Associated Herpesvirus (KSHV) that contains 26 and 22 short repeats each with 3-4 G’s each (9) as well as the C9orf72 gene locus in which expansion of 5’-GGGCC-3’ repeats from 2-24 in the genome to sometimes several thousand copies is the cause of the most common form of familial amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (10). G-quadruplexes form when runs of two or more guanine residues in a DNA stack on each other in a square planar structure stabilized by Hoogsteen base pairing (8,11,12). They are further stabilized by certain cations, in particular K+ (13,14).

The identification of telomeric G-quadruplexes in vitro was first reported in 1989 (15,16). The existence of G-quadruplexes in vivo was initially recognized by immunohistochemistry (IHC) in the telomeric DNA of the ciliate Stylonychia lemnæae (17,18) and later throughout the human genome (19). Only a portion of the human cellular G-quadruplex IHC signal was detected at telomeres consistent with the presence of multiple and likely different repeat sequences throughout the genome capable of forming G quadruplexes (19).

There has been growing interest in the potential role of G quadruplex formation at the ssDNA termini of mammalian and many other eukaryotic telomeres in which the G-rich strand extends beyond the double strand DNA (dsDNA)-ssDNA junction as a 3’-ss overhang. In humans, this ssDNA extension is ~150-300 nt in length (7) and studies have pointed to the formation of G quadruplexes in the overhang and their role in inhibiting telomerase as well as regulating binding and action by the shelterins (20).

A 24 nt ssDNA consisting of (3’-TTAGGG-5’)4 has been crystalized as a model for the telomeric overhang. In the resulting structure the DNA is arranged into a flat disk 0.6 nm in height and 4 nm on a side with the DNA adopting a parallel G-quadruplex conformation, and with TTA triplets looping out in a propeller fashion (21). However, NMR studies and other solution approaches have argued that the 24 nt units exist in hybrid or anti-parallel conformations suggesting that the parallel arrangement may be promoted by crystallization (22). Potential condensation of G-quadruplex folds into larger structures have been proposed but not experimentally demonstrated (23).

Telomeric repeats also exist in extremely long nucleic acid chains in the form of G-rich transcripts of telomeres termed TERRA (5’-UUAGGG-3’)n. In human cells TERRA molecules of nearly 9000 nt in length have been observed (24,25). We previously demonstrated that
TERRA is arranged into chains of 24 nt particles joined by 3 nt linkers and provided evidence that these particles are stabilized by G-quadruplex folds (26). Thus, a 24 nt particle stabilized by G-quadruplex formation may be shared by both the telomeric G-rich ssDNA and RNA forms. Moreover some TERRA RNA remains present in telomeric R-loops following transcription in many organisms (27). Telomeric R-loops have been shown to promote homologous recombination (HR) and are prevalent in a significant number of cell lines and tumors exhibiting the Alternative Lengthening of Telomeres (ALT) phenotype (27).

It is important that we understand the kinds of macroscopic folding that ssDNA segments anywhere from 200 to 2000 nt or more in length can adopt at sites of transcription, recombination, or uncoupled replication forks. To do so requires generating very long ssDNA substrates so that any effect of DNA ends on the internal folding is minimized. In this study, we generated very long ssDNA (up to 20,000 nt) composed of the G-strand telomeric repeat (3’-TTAGGG-5’), the C-strand complement repeat (3’-AATCCC-5’) and a mutant G-strand repeat (3’-TTAGTG-5’). We then interrogated their structure using EM and single molecule magnetic tweezers (smMT) force extension analysis, two methods capable of providing structural information about very large nucleic acid molecules. We demonstrate that the G-strand telomeric ssDNA spontaneously condenses into a thick bead-like filament as imaged by EM and resolved by step-wise elongation following force extension analysis by smMT. Moreover, we show that switchable ATP-dependent RAD51 ssDNA binding could be used to probe the structure of these higher-order G-strand configurations.

RESULTS
Preparation of Long G-Strand Telomeric Single Stranded DNA.
Tandem copies of ssDNA containing the human G-strand (5’-TTAGGG-3’)_n, repeat, the complementary C-strand (5’-CCCTAA-3’)_n, repeat and a mutant repeat disrupting the three-G run (5’-TTAGTG-3’)_n were generated using a φ29 DNA polymerase rolling circle replication scheme from a mini-circle template as previously described (Supplementary Figure S1a) (28). The size distribution of the ssDNAs was determined by alkaline gel electrophoresis and displayed a broad distribution from 100 nt to >20,000 nt (Supplementary Figure S1b).

G-strand Single Stranded DNA Contains G-Quadruplex Structures.
Numerous studies have indicated that telomeric G-strand ssDNA folds into G-quadruplex structures (15, 16, 29). CD analysis is frequently used to demonstrate the presence of G-quadruplex structures within DNA (30). However, CD studies require high molar concentrations of a single DNA species and the long ssDNAs prepared by the rolling circle method contain a mixture of many different lengths even following alkaline agarose gel isolation and purification, resulting in very low molar concentrations of any individual species. Spectra obtained and further discussion of the limitations is presented in the Supplementary Material and Supplementary Figure S2.

As an alternative we utilized structure-specific G-quadruplex Fluorescence Intercalator Displacement (G4-FID) analysis to probe the underlying structure of the long ssDNA generated by rolling circle replication (Figure 1a; (31)). The G4-FID assay was developed from similar assays of double stranded DNA (dsDNA). When thiazole orange (TO) binds DNA, it produces a strong fluorescence signal with high quantum yield while remaining non-fluorescent in the absence of DNA. Displacement from DNA may be monitored as a decrease in fluorescence. This assay is now used widely to determine selectivity of a candidate compound through its ability to displace TO from quadruplex DNA (31-33). We used the 3,6,9-trisubstituted acridine ligand BRACO-19, which was
designed by molecular modeling as a G-quadruplex binding-specific candidate molecule. Previous work has shown that BRACO-19 inhibits telomerase activity, which results in telomere shortening and end-to-end chromosomal fusions in cancer cells as a consequence of quadruplex stabilization at the ssDNA overhang (34,35). Indeed a crystal structure of the human telomeric G-quadruplex and BRACO-19 has been solved (36) and resembles the native unimolecular and bimolecular human telomeric quadruplexes structure (21) bound by TMPyP4; a well-known porphyrin based G-quadruplex binder (37). BRACO-19 was used earlier by another group as candidate G-quadruplex binder for G4-FID assays (33). We introduced increasing BRACO-19 to long G-strand ssDNA which resulted in reduced TO fluorescence and increased the calculated TO displacement (Figure 1b, blue triangle; EC$_{50}$ = 0.44 μM). Moreover, the TO displacement of equivalent concentrations of the smaller (5'-TTAGGG-3')$_4$ (G24; Figure 1b, black square; EC$_{50}$ = 0.43 μM) and (5'-TTAGGG-3')$_{20}$ (G120; Figure 1b, red circle; EC$_{50}$ = 0.45 μM) G-quadruplex forming oligonucleotides by BRACO-19 appeared nearly identical to the long G-strand DNA. Importantly, we found that BRACO-19 displaced TO from the G-strand ssDNA (Figure 1c, black square; EC$_{50}$ = 0.42 μM) but not the complementary C-strand (Figure 1c, red circle; EC$_{50}$ = 7.3 μM) or the mutant G-strand ssDNA that disrupts the G-quadruplex repeat (Figure 1c, blue triangle; EC$_{50}$ = 7.2 μM). These observations are consistent with the conclusion that the long G-strand ssDNA forms G-quadruplex structures similar to well-defined G-quadruplex forming oligonucleotides that are known to bind the G-quadruplex specific ligand BRACO-19. In contrast, long C-strand and mutant G-strand ssDNA, which lack the G-quadruplex structure do not bind BRACO-19.  

The G-strand, C-strand, and mutant G-strand ssDNA's were prepared for EM by direct adsorption to thin carbon supports followed by tungsten metal shadow casting. The C-strand DNA appeared thin and lacked any discernable organization, typical of phage ssDNAs prepared by these methods (Figure 2a). The mutant G-strand DNA (Figure 2b) was also disorganized although these molecules appeared to display more compaction that the C-strand DNA. We determined the width of tungsten metal shadow casting for the C-strand (1.8 ± 0.6 nm, N = 268) and found that it appeared approximately half the width of M13 dsDNA (2.9 ± 0.5 nm, s.e.m., N = 125) (Supplementary Note S2). The results suggest that in these preparations both the C-strand ssDNA and M13 RF dsDNA were coated with 0.9 nm of tungsten metal in order to account for their known ~1 nm and 2 nm width, respectively.

In the EM images the G-strand ssDNA was remarkably different and consisted of thick condensed fibers with a bead-like appearance that were often linked by thin filaments similar in thickness to the C-strand DNA (Figure 2c,d, Supplementary Figure S3). The diameter of these bead-like particles from multiple preparations revealed two size classes (Figure 3a). The larger bead-like particles (9.4 ± 1.9 nm, s.d.; Figure 2c,d, large arrowheads) when corrected for tungsten metal deposition had an estimated mean diameter of 8.5 nm, while the smaller bead-like particles (6.1 ± 1.5 nm, s.d.; Figure 2c,d, small arrowheads) when corrected for tungsten metal deposition had an estimated mean diameter of 5.2 nm. In numerous cases the larger particles appeared conjoined in a way that appeared to suggest an association of two smaller bead-like particles might generate the larger ones (see also Supplementary Figure S3; Supplementary Note S2). In all of the fields of the G-strand ssDNA some molecules were less organized suggesting that either other conformations may be possible or that the process of
forming the beaded structures may be incomplete.

The G-strand ssDNA was subjected to alkaline agarose gel electrophoresis and a region corresponding to 1-1.5 knt was isolated, neutralized and purified. This size-selected G-strand DNA (ave. 1.25 knt) was prepared for EM in the presence of K⁺ cation and the length and number of beads were determined for molecules that appeared fully saturated with mostly larger bead-like particles and no significant linker DNA (Figure 3b, left). Compared to the calculated length of 875 nm for a random sequence 1.25 knt ssDNA (0.7 nm/nt for ssDNA) (38), the mean length of the saturated G-strand ssDNA (69.5 ± 18.9 nm, s.d.; N = 91; Figure 3b, right) appeared compacted 12.6-fold. By dividing the length of each individual molecule by its distinct number of beads we determined the mean length per bead from the binned values (13.7 ± 1.9 nm/bead, s.d.; N = 91; Figure 3c); equivalent to 5 beads per 1250 nt. Together, these observations suggest that there are approximately 250 nt of G-strand ssDNA per large bead particle.

In the EM experiments described above, the DNA was adsorbed to the thin carbon foils in a buffer containing 2.5 mM spermidine (Experimental Procedures) followed by dehydration through water-ethanol solutions and air-drying. While this approach has been found to faithfully reveal structures in DNA in past studies, and provides the highest density of material on the substrate, we also mounted the DNA using a new method being developed in this laboratory. Here the DNA was adsorbed to the supports in a low salt buffer containing no divalent metals or spermidine followed by very rapid freezing, freeze-drying and finally rotary metal shadow casting. While the images were not as crisp due to residual salt remaining after freeze-drying, the basic beaded appearance of the ssDNA was the same as when prepared by the procedure used in the experiments above (Supplementary Figure S4).

The G-Strand Single-Stranded DNA is Resistant to Extension Force.

To further examine the properties of the G-strand ssDNA we developed a smMT system (39-41) in which one end of the G-strand ssDNA was attached to an engineered flow cell surface via a biotin-NeutrAvidin linkage, while the other end of the DNA was attached to a superparamagnetic (SPM) bead via a digoxygenin-antidigoxigenin linkage (42). Controlled positioning of permanent neodymium magnets allowed the application of a well-defined force that results in DNA extension, which was monitored by the location of the SPM bead relative to the flow-cell surface (39,41,43).

Alkaline agarose gel electrophoresis was used to size-select ~5 knt length C-strand, G-strand and mutant G-strand ssDNA in which the ligated rolling-circle contained biotin and the 5’-thymine residue of the primer oligonucleotide contained digoxygenin (Supplementary Figure S1). The C-strand and mutant G-strand ssDNA's displayed a force extension curve consistent with a freely jointed chain model that is characteristic for ssDNA (Figure 4a, black squares with red curve; Supplementary Figure S5) (44). In contrast, a 5 kb dsDNA displayed a force extension that is consistent with a worm-like chain model as previously described (Figure 4a, red squares with blue curve; Supplementary Figure S5) (45). Remarkably, the G-strand ssDNA extension stalled at ~600 nm that resulted in a force-extension curve which adopted a sigmoidal character (Figure 4a, blue squares with turquiose curve; Supplementary Figure S5). These results are inconsistent with a mechanical extension that is caused by the natural ordering of a disordered polymer molecule under force (39), but rather implies a compacted ordered DNA structure that is relatively resistant to forces below 12 pN (Figure 4a; Supplementary Figure S5). Interestingly, at forces above 5 pN (Figure 4a,b, red box; Supplementary Figure S5) we observed step-wise extension of the G-strand ssDNA rather
than the continuous smooth extension exhibited by the mutant G-strand (Figure 4b,c; Supplementary Figure S5).

A large positive change in the second-derivative of the extension versus time plot (Figure 4c) was used to identify 434 steps from 62 individual DNA molecules (Figure 4d). When binned to the resolution of the smMT instrument (10 nm) we observed a distribution of steps that appeared to contain multiple embedded peaks (Figure 4d). Employing coefficient of determination (46) and root-mean-squared goodness of fit (46) analysis appeared to indicate six embedded Gaussian peaks that when smoothed describe the binned histogram data as corresponding to six step-release sizes (Figure 4d; Supplementary Figure S6).

Interestingly, we observed a weak positive trend between the step size and the introduced force, with the largest steps occurring at higher forces (Supplementary Figure S7). It is important to note that even at very slow load rates the release of folded G-quadruplex structure only begins to occur at forces above 20 pN and peaks at 35-55 pN, depending on the force spectroscopy method (47,48). Together, these results seem to imply that the application of magnetic force results in the step-wise release of higher-order structure(s) within the G-strand ssDNA, while maintaining the intrinsic array of G-quadruplex folds.

**Modeling**

Several models have been proposed for the structure of the 3'-ssDNA overhang at the terminus of mammalian telomeres, which are largely based on parallel or hybrid G-quadruplex structures and result in somewhat different outcomes (23). For example, the crystal structure of a 22 nt d[AG₃(T₂AG₃)₃] ssDNA (21) revealed a flattened disk 0.61 nm in height and 4.2 nm in diameter with the four unpaired TTA triplet nucleotides extending out to the edges of the disk similar to propellers. Other crystal structures have been derived from similar short DNAs bound with different ligands, where with the ligand dominates the outcome (23).

Multimeric G-quadruplex structures are largely based on beads-on-a-string models (29), where multiple G-quadruplex units formed from (3'-TTAGGG-5')₄-20 are arranged into individual thermodynamically unique beads-on-a-string that do not interact with one another, can move freely and are only restricted by their TTA linker. These models are founded on studies of limited length DNAs (120 nts or less) and do not address the possibility of a variation in the length of intervening repeat ssDNA or interactions between G-quadruplex units via stacking interactions of the G-tetrad cores (23). Perhaps more importantly there are several types of G-quadruplex folds as well as suggestions that these different folds may exist together under identical conditions (49). Nevertheless, once folded all G-quadruplex structure appear exceedingly stable (50).

To understand the distribution of step-sizes in the force extension analysis we performed G-quadruplex folding simulations with variable lengths of G-strand ssDNA (Supplementary Figure S8). In these simulations, a G-quadruplex may fold randomly along the length of an ssDNA between four contiguous telomeric repeats. The type of G-quadruplex fold is irrelevant for these simulations, although we assume that once formed these folds are stable and no subsequent remodeling may occur. These random folding events result in an array of G-quadruplex structures along the G-strand ssDNA that are separated by 0, 1, 2, or 3 repeats with distinct probabilities (Supplementary Figure S8).

We determined that a G-quadruplex embedded in a 1 knt or 6 knt G-strand ssDNA is on average separated from the adjacent G-quadruplex by one complete telomeric repeat (Supplementary Figure S8). When subjected to smMT at forces below those that disrupt the G-quartet structure, a single G-quadruplex may be pulled from two different corners that adds different lengths depend on the type of fold (51). This corner-to-corner extension would
account for 0.6-1.5 nm that with a free TTA tail (2.1 nm) and on average one telomeric repeat (4.2 nm) totals of 6.9-7.8 nm. The projected force extension peaks appear to be tetramer multiples of the ~7.3 nm unit (calculated: 29 nm, 58 nm, 87 nm 117 nm and 146 nm; observed 33 nm, 55 nm, 77 nm 117 nm and 142 nm; Figure 4d). The extra 99 nm (or 77 nm) extension peaks might reflect intermediate extensions and/or over-fitting of multiple Gaussian curves.

A plausible model for the higher-order structures exploits the possibility that random condensation along G-strand DNA results in G-quadruplexes that are separated by, on average, one additional G-strand repeat (Figure 5a). These intervening repeats contain guanine nucleotides that could form Hoogsteen base pairing with nearby intervening G-repeats that are not folded into a G-quadruplex (11, 14). It is our hypothesis that four intervening repeats condensed into a G-residue pseudo-fold, which is the basis for stable higher-order packing (Figure 5b, dark blue box). In this scenario four intervening repeats could come from any combination of 1-3 repeats between two embedded G-quadruplexes linked with the additional necessary nearby intervening repeats such that a total of four repeats are condensed into a G-residue pseudo-fold. The extension of these G-residue pseudo-fold higher-order structures under the force conditions utilized in our studies would leave the bona fide G-quadruplexes intact while stretching the terminal TTA linkers as well as the intervening G-strand repeats on average ~29 nm (Figure 5c). Such a G-residue pseudo-fold formed between intervening repeats could account for the increased force required to extend the G-strand ssDNA that is significantly more than random-coil ssDNA but less than a true G-quadruplex fold (Figure 4d). Importantly, this model predicts the G-residue pseudo-folds formed between multiple intervening repeats is more exposed than the G-strand ssDNA folded into a compact G-quadruplex.

RAD51 systematically alters higher order G-quadruplex associations

To further probe the structural characteristics, we examined the force extension properties of G-strand ssDNA in the presence of the central HR component RAD51. Unlike strong ssDNA binding proteins that singularly disrupt G-quadruplex folds formed in ssDNA (52), the RAD51 protein displays a switchable weak ssDNA binding in the absence of ATP (or in the presence of ADP), that is transformed into a robust nucleoprotein filament in the presence of ATP which firmly coats a ssDNA (53). This cycle of RAD51 ATP binding and hydrolysis has been connected to nucleoprotein filament reordering and turnover during the HR homology search (54) and may play a role in ALT telomere maintenance (55). Importantly, the formation of an ATP-bound RAD51 nucleoprotein filament appears capable of generating sufficient binding forces to provoke nucleosome disassembly in model chromatin substrates (56).

We first determined that multiple cycles of magnetic extension and release of a single G-strand ssDNA resulted in largely identical force extension curves (Figure 6a; Supplementary Figure S9a). This observation strongly suggests that the G-strand ssDNA spontaneously re-folds into comparable stable higher order structures following release from magnetic extension. When RAD51 is included in the force extension analysis without ATP (-ATP) the curves gradually flatten following multiple extension and release cycles, demonstrating significantly more extension at low forces (Figure 6b; Supplementary Figure S9b). Combined with the G-quadruplex modeling (see Modeling above and Supplementary Figure S8), we interpret these results to imply that RAD51 increasingly binds to intervening ssDNA repeat(s) embedded between G-quadruplex folds; eventually inhibiting spontaneous condensation into the G-residue pseudo-folds that constitute the higher order structures. We note that the weak and reversible ssDNA binding activity of RAD51
(-ATP) does not appear to fully inhibit higher order structure re-folding and results in a DNA-protein complex that does not exhibit ssDNA or dsDNA force extension properties (Figure 6b; Supplementary Figure S9b). In contrast, RAD51 in the presence of ATP (+ATP) appears to rapidly form a stable nucleoprotein filament that extends longer than a similar length dsDNA at low forces and displays no intermediate force extension curves regardless of the number of extensions (Figure 6c; Supplementary Figure S9c). The longer extension of the RAD51 nucleoprotein filament is consistent with historical studies that demonstrated a helical ATP-bound RAD51 nucleoprotein filament extends an ssDNA 1.5x the length of a corresponding dsDNA (57,58).

To confirm the hypothesis that RAD51 form a stable filament that fully extends the G-strand ssDNA we examined the RAD51 nucleoprotein filament in the presence of ATP by EM following negative staining. We observed fields of extended RAD51 nucleoprotein filaments with a helical pitch of 10.7 nm (n= 59 measurements) in agreement with the previous studies (Figure 7). Similar filaments were not present when G-strand ssDNA was incubated with RAD51 in the absence of ATP (not shown). We conclude that RAD51 forms a stable nucleoprotein filament in the presence of ATP that fully extends the G-strand ssDNA, unraveling both the embedded G-quadruplex folds and the higher-order G-residue pseudo-fold structures.

DISCUSSION

Extending the structural studies of short (~20-150 nt) G-strand repeat ssDNAs to the analysis of much longer G-strand repeat ssDNAs has been hampered by the lack of well-defined substrates. To overcome this technical issue, we further developed a rolling circle replication system as previously described (59,60) and adapted it for generating long ssDNA containing distinct repeating nucleotide sequences. This method allowed us to produce G-strand, C-strand and mutated G-strand ssDNAs containing different forms of the human telomeric repeat sequence ranging in size from 100 to >20,000 nt (Supplementary Figure S1).

G4-FID analysis confirmed the presence of G-quadruplex structures within the long human G-strand ssDNA (5'-TTAGGG-3')ₙ, but not in the C-strand ssDNA (5'-AATCCC-3')ₙ or a mutant G-strand ssDNA (5'-TTAGTG-3')ₙ that interrupts the contiguous G-triplet within the telomere repeat. Using EM we showed that very long human telomeric G-strand ssDNA spontaneously condenses into chains of large discrete bead-like particles with 5 nm and 8 nm diameter. These bead-like particles compact the ssDNA nearly 12 fold in length. Using size-selected DNA we determined that the larger beads contain ~240 nt and inferred the smaller beads contained half this amount (~120 nt).

Unlike EM, which requires a final dehydration step, smMT studies may be performed in solution at physiological ionic strength and with predominantly intracellular K⁺ cation. Force extension analysis indicated several discrete G-strand ssDNA elongation step sizes. Because it requires extension forces >20 pN to unravel a folded 24 nt G-quadruplex (47,48), we interpreted these results to indicate that the extension steps represented fundamental units of higher-order interaction structures. The weak positive correlation between step size and extension force (Supplementary Figure S7) suggests that the higher-order structures are intrinsically more stable the larger they become. In EM fields of the compacted G-strand ssDNA we noted some particles appeared less organized than others. This might suggest that in addition to the principal structures other arrangements may be possible that might contribute to the force-extension "noise" that extends outside the baseline 10 nm resolution of the smMT instrument. Taken together these observations represent the first physical evidence for intra-molecular higher-order structures formed by G-strand ssDNA containing embedded G-
quadruplexes. These findings appear broadly applicable for the long repetitive blocks of G-rich sequences present in many genomes including the human genome.

We calculated that the major tetramer and octamer higher order structures observed as initial step sizes by smMT would on average contain ~120 nt (four G-quadruplexes) and ~240 nt (eight G-quadruplexes and eight extra repeat loops), respectively. The amount of the octamer higher order structure DNA content is remarkably consistent with the calculated amount of ssDNA associated with the large EM beads (247 nt), and by extension the DNA content of the smaller beads that appear approximately half the size. These two initial smMT extension steps account for 63% of the total events, which appears consistent with the two major bead sizes observed by EM analysis.

We developed a model based on the stochastic formation of G-quadruplexes, followed by condensation of intervening G-repeats (Figure 5). An inter-molecular side-by-side pairing of a G-repeat ssDNA, termed G-wires, has been described that utilized small oligonucleotides (5'-GGGGTTGGGG-3') (61). In theory, G-wires might provide a plausible structure for the association(s) of intervening telomeric ssDNA repeats. However, we note that the inter-molecular G-quadruplex folds predicted by G-wires should display comparable stability to force-extension as bona fide intra-molecular G-quadruplexes. A less robust interaction between the intervening repeat G-residue pseudo-folds appears more consistent with our studies, which showed both magnetic force and the weak ssDNA binding activity of RAD51 (-ATP) were capable of disrupting the higher-order structures. For example, we interpret the observation that RAD51 (-ATP) increasingly inhibits condensation following multiple release of magnetic force as consistent with increased shielding of intervening repeat ssDNA by weakly bound RAD51; supporting the intervening repeat G-residue pseudo-fold condensation model (Figure 5).

We performed extensive nuclease digestion analysis of the long G-strand ssDNA exploiting multiple nuclease with the goal of isolating higher order G-strand structures (not shown). However, under the best conditions we observed a 24 nt band consistent with a protected G-quadruplex along with a smear of higher molecular weight ssDNA. We concluded that the higher-order G-strand structures are only partially resistant to nuclease, which also appears consistent with exposed features of intervening repeat G-residue interactions as proposed in our model (Figure 5).

We calculated the probability (P) that a higher-order G-strand structure might condense within various lengths of human telomere 3'-ssDNA by assuming that a minimum of four individual ssDNA G-strand repeats separated by at least three G-quadruplex folds are required to form a stable particle (Supplementary Figure S8e; Methods) (62). This analysis suggested that a significant fraction of 100 nt (P = 0.13) and 200 nt (P = 0.32) human telomeric G-strand ssDNA may stochastic fold into G-quadruplexes leaving at least four intervening G-repeats, which could condense into a G-residue pseudo-fold. This probability dramatically increased with increasing ssDNA length (Supplementary Figure S8e). Moreover, these estimates likely represent a minimum since we do not include the possibility that adjacent ssDNA repeats might nucleate into a G-residue pseudo-fold or consider any potential phasing by the telomere ssDNA-dsDNA junction. Taken as a whole, these studies suggest that in the absence of other factors, higher-order G-strand structures may form naturally on a 3'-ssDNA (63,64).

We consider the possibility that a higher order telomeric G-strand structure could serve as a molecular switch that might inhibit or facilitate the interactions of telomeric protein factors such as the shelterin protein complex or telomerase. Compaction into these large structures
could aid in shielding the telomere from double strand break (DSB) sensing and repair factors prior to binding by telomeric protein components (65), or they might act as a roadblock for telomerase mediated telomere elongation (13). Cech and colleagues have examined a 144 nt (3’T- TTAGGG-5’) ssDNA bound by hPOT1 and found highly compact DNA-protein particles, suggesting that POT1 is likely to disrupt any higher order G-strand condensation. Indeed POT1 and its partner TPP1 appear to overcome G-quadruplex roadblocks during telomere metabolic processes (66).

The self-condensation of long segments of ssDNA containing runs of G’s could have significant global consequences for replication, recombination and repair. As noted above, if the G-rich strand at the C9orf72 locus (C2G4)n assumed a higher order conformation, this could not only induce replication fork stalling, but in itself might lead to further expansion by mechanisms related to replication restart, recombination or repair. Indeed, the expanded repeat has been shown to cause replication fork stalling in vitro (67). The KSHV replication origin contains ~260 bp of G-rich repeats on one strand. Were this G-rich strand to condense into a higher order structure, this would likely generate a severe bend or knot at the KSHV replication origin with the C-rich strand present in a more open unstructured state. The net result would likely be a complex scaffold with very different protein binding properties than the linear dsDNA.

It will be of interest to generate other long ssDNAs including ones containing the G-rich C9orf72 repeat and the KSHV origin repeats and examine their architecture using the biophysical methods described here.

EXPERIMENTAL PROCEDURES

Rolling Circle Replication

Telomeric ssDNA circles were synthesized using oligonucleotides with either wild type G-strand, mutant G-strand, or C-strand sequences (see below) containing 20 telomeric repeats (IDT, Coralville, IA) as described earlier (28). In brief, oligonucleotides were ligated into circles using CircLigase™ and any remaining linear oligonucleotides were digested with ExoI and ExoIII (New England Biolabs Inc., Ipswich, MA) according to the manufacturer’s recommendations (Epicentre Biotechnologies, Madison, WI). Circular products were confirmed by 7% denaturing PAGE.

For size selection of long G-strand DNA we used alkaline agarose gel electrophoresis. In brief, 0.8% alkaline agarose gels were prepared in alkaline buffer (10X buffer contains 10N NaOH and 0.5M EDTA pH-8.0). The sample DNA was digested with S1 nuclease and mixed with alkaline gel loading buffer. After electrophoresis, the gel was soaked in neutralizing solution (containing 1M Tris-HCl and 1.5M NaCl), stained with SYBR® Green I Nucleic Acid Gel Stain (Invitrogen) followed by isolation of target DNA length using a QIAquick gel extraction kit (Qiagen).

G-Quadruplex Fluorescent Intercalator Displacement (G4-FID)

A constant temperature (20°C) SPEX Fluorolog-3 (Horiba) spectrofluorometer with thermostated cell holders (3 ml) was used to perform G4-FID studies in 10 mM Li-cacodylate buffer (pH 7.3) and 100 mM KCl. Briefly, 0.25 µM pre-folded DNA in Li-cacodylate-KCl buffer was mixed with thiazole orange (TO) (0.50 µM). Each ligand addition step (from 0.5 to 10 equivalents) was followed by a 3 min equilibration period after which the fluorescence spectrum was recorded. The percent displacement was calculated from the fluorescence area (FA, 510–750 nm, λex = 501 nm), using the following equation: TO displacement (%) = 100−[(FA/FA0)×100], where FA0 is the fluorescence of TO bound to DNA without added ligand. The TO displacement (%)
was plotted as a function of the concentration of added ligand.

**Electron Microscopy**

DNA was adsorbed onto grid supports covered with a thin glow discharge treated carbon film for 30 sec to 1 min in the presence of the buffer present with the DNA, or a buffer containing 2.5 mM spermidine, 10 mM Tris-HCl, 50 mM KCl, 75 mM NaCl and 1 mM MgCl₂ (pH 7.5). Samples were washed in water followed by a series of ethanol dehydration steps, air-drying, and rotary shadow casting with tungsten at 1×10⁶ torr (68). Samples were visualized using a Tecnai 12 TEM (FEI Inc., Hillsboro, OR) at 40 kV and images were collected with a Gatan Orius CCD camera (Gatan Inc., Pleasanton, CA) with Digital Micrograph supporting software (Gatan Inc., Pleasanton, CA). Dimensions of telomeric DNA filaments were measured from digital micrographs using Digital Micrograph (Gatan Inc., Pleasanton, CA) and ImageJ (NIH, Bethesda, MD). Negative staining of the G-rich ssDNA bound by human RAD51 (purified as previously described (56)) was carried out by incubating 50 ng of G-rich ssDNA with 1500 ng of RAD 51 in a buffer containing 20 mM HEPES (pH 7.5), 10% glycerol, 0.5 mM DTT, and 2 mM MgCl₂ for 30 min at 37 degrees in the presence or absence of 2.5 mM ATP. Drops of the sample were adsorbed to glow charged thin carbon supports for 3 min followed by washed with 2% uranyl acetate, air drying and imaging in a Tecnai 12 as above at 80 kV. Images for publication were arranged and contrast optimized using Adobe Photoshop CS5 (Adobe Systems, San Jose, CA).

**Magnetic Tweezers Preparation**

Flow cells were engineered with glass cover slides affixed with double-sided tape to an aluminum foundation that maximized SPM bead imaging. Telomeric C-strand, G-strand or mutant G-strand ssDNA (30 pM final) was mixed with bead formation buffer (50 mM Tris-Cl (pH-7.6), 100 mM KCl). The combined sample was boiled for 10 minutes followed by snap chilling on ice for 1-2 hours. Prior to attachment, the glass slides were treated with (3-Aminopropyl) triethoxysilane followed by a 1:100 mixture of Biotin-PEG SVA to mPEG-SVA (Invitrogen). Neutravidin (500 µm, Invitrogen) was injected in the flow cell at a rate of 8 µl min⁻¹, followed by the ssDNA. Tosyl activated M-280 SPM Dynabeads (ThermoFisher Scientific) were coated with anti-digoxigenin antibodies (Roche). Stock beads were removed and resuspended in a 0.1 M Borate, pH 9.5 (Buffer A). Beads were then resuspended in a mixture of Buffer A and anti-dig at a 20 µg: 1 mg, antibody: beads ratio. The beads were incubated for 12-17 hours at 37°C with slow tilt rotation. 3M Ammonium Sulphate in Buffer A (Buffer B) was mixed into the beads. The beads were then resuspended in 10 mM Na/K-Phosphate (pH 7.4), 140 mM NaCl/KCl (PBS) with 0.5% (w/v) Acetylated BSA (Sigma) (Buffer C). After incubation at 37°C, 1 hr, the coated beads were resuspended in PBS (pH 7.4) with 0.1% (w/v) Acetylated-BSA (Buffer D). Buffer D was removed and then added to reach the desired bead concentration (2×10⁹ beads mL⁻¹).

Anti-digoxigenin coated beads were washed with Running Buffer [50 mM Tris-Cl (pH-7.6), 100 mM KCl, 200 µg µL⁻¹ Acetylated BSA, 0.0025% Tween-20 (Amresco)], and injected into the flow cell containing the ssDNA that was bound to the surface via a biotin-neutravidin linkage; at 8 µl min⁻¹ while agitating the system. The bound ssDNA was washed extensively with running buffer to remove free SPM beads prior to analysis. Force extension measurements used four 1 cm³ rare earth magnets (Neodymium, Magcraft). The SPM beads were imaged using a 530 nm LED lamp (Thorlabs), a 100X Olympus oil immersion objective and images collected on a 1024 x 1024 pixel charge coupled device (CCD) camera (Grasshopper Express 1.0 MP Mono FireWire 1394b) at a frame rate of 70 ms.
The human RAD51 protein was purified and stored as previously described (52). Force extension analysis was performed in Running Buffer additionally containing 2 mM MgCl$_2$, 100 µM Dithiothreitol (DTT) and human RAD51 (500 nM) with or without 1 mM ATP (Roche).

**Magnetic Tweezers Data Analysis**

SPM bead analysis was performed with the 3D bead tracking software Video Spot Tracker (CISMM at UNC-CH). Displacement events were determined using the software Edge Detector (CISMM at UNC-CH) and MATLAB and Statistics Toolbox Release 2014b (The MathWorks, Inc., Natick, MA). Edge Detector was edited to avoid scoring negative displacement as these can bias the value of large positive displacement. Data was smoothed with the Savitzky-Golay filter$^3$ and an average window size of 20 points using Origin software (OriginLabs, Northhampton, MA). Distributions for displacement events were binned to the resolution size of step size (~10 nm) and

\[
\text{[size of bins} = \frac{\text{Max} - \text{Min}}{\text{Number of Bins} - 0.5} \text{], where:}
\]

Max = maximum value of events; Min = minimum value of events; Bin Start = Min; Bin End = (bin size x number of bins). The events were fit to multiple Gaussian distributions to determine several different observed displacement changes. The coefficient of determination for different numbers of peaks was used to select the appropriate number of Gaussians to fit. All fittings were done with Origin software. Force and extension values were determined for various magnet positions above the flow cell surface. Force calculations were determined by measuring the bead’s fluctuations transverse to the direction of stretching and equating the fluctuation to the equipartition theorem (43). Double strand and mutant G-strand DNA curves were fit to the freely jointed chain model for DNA nonlinear elasticity (56).

### Simulation of Multiple Quadruplex Formation

Define an $n$-dimensional array $\bar{a}$, composed of $a_i$, $i = 1, \ldots, n$ where $a_i = 0$ for all $i$.

Define Algorithm A:

Create $n$ random numbers in an $n$-dimensional array $\bar{u}$, composed of $u_i$, $i = 1, \ldots, n$ where $u_i \sim \text{iid } U([0,1])$.

Sort the array $S(\bar{u})$, to create ordered uniformly random numbers $u_{[i]}$, $i = 1, \ldots, n$ where $u_{[i]}$ is the $i^{th}$ sorted value of $\bar{u}$. Then let $I(u_{[i]})$ be the original index of the $i^{th}$ sorted value of $\bar{u}$.

\[
\text{for } i = 1 \text{ [}
\begin{array}{l}
\text{if } \sum_{j=0}^{3} a_{I(u_{[i]})+j} \text{ is defined and } = 0 \text{ then do;}
\end{array}
\]\[1\]
\[
\text{for } j = 0 \text{ [}
\begin{array}{l}
a_{I(u_{[i]})+j} = 1
\end{array}
\]\[1\]
\[
\text{end;}
\]
\[
\text{end;}
\]

Algorithm A results in $\bar{a}$ having a series or set of blocks $B$ where:

\[
b \in B = \{ \{ I(u_{[i]}), I(u_{[i]})+1, I(u_{[i]})+2, I(u_{[i]})+3 \}, \text{ for all } i = 1, \ldots, n \text{ satisfying } \sum_{j=0}^{3} a_{I(u_{[i]})+j} = 0 \text{ after potentially reassigning values for } a_i \text{'s through } (i-1) \text{ previous iterations.}
\]

Note that this implies $a_{I(u_{[i]})+j} = 1$ for $j = 0, 1, 2, 3$ for each element $b \in B$ after Algorithm A.

That is, each element $b$ of $B$ is a set or block of four consecutive indices denoting consecutive positions in the array $\bar{a}$ each having a value of 1 with gaps between blocks having size 0, 1, 2, or 3, which have consecutive values of 0.

Let $g_k$ be the size of the $k^{th}$ “gap” between the block $b_k$ and $b_{k+1}$, $g_k \in \{0,1,2,3\}$ for all $k$.
= 1, ..., m where m ~ M is a random number representing the total number of gaps that is dependent on Algorithm A and each iteration of Algorithm A. The distribution of M over iterations of algorithm A is potentially obscure and may not have a simple representation. The frequency of each gap length within an iteration of Algorithm A,

\[ i.e., N(g_k = 0 ) = N_0, N(g_k = 1 ) = N_1, N(g_k = 2 ) = N_2, N(g_k = 3 ) = N_3 \]

are also random numbers from different distributions that are dependent on n. By definition,

\[ N_0 + N_1 + N_2 + N_3 = m \]

**Calculating the Probability of Higher-Order G-strand Structure Formation**

A solution to the likelihood that a human telomere G-strand repeat sequence (TTAGGG\textsubscript{n}) supports an environment compatible with higher-order structure formation begins by defining a condition where four consecutive repeats (TTAGGG\textsubscript{4}) fold into a G-quadruplex, followed by 1-3 repeats (TTAGGG\textsubscript{1-3}; termed: a gap). This chain of G-strand repeats can be thought of as a “unit” containing a G-quadruplex fold with adjacent gap, surrounded by G-quadruplex folds containing no gaps. Structural analysis has suggested that three gaps with an adjacent G-quadruplex (3 units) plus one additional gap of 1-3 G-strand repeats are minimally required to form a higher-order structure. Thus, we can reframe the probability of a nucleotide sequence having conditions compatible with higher-order structure formation as the probability that at least four successive gaps will occur within the G-strand sequence. If we term developing a single gap as a “success”, then Feller (62) described a simple method for determining the probability of at least \( r \) consecutive successes (gaps) in \( n \) Bernoulli trials (total gaps):

\[
P \left( \text{r consecutive successes} \right) \sim 1 - \frac{1 - px}{(r+1-rx)q} * \frac{1}{x^{n+1}} \quad \text{(eq. 1)}
\]

where, \( p \) is the probability of success of a single gap, \( q = 1 - p \), and \( x \) is the real root:

\[ 1 - x + q * p^r * x^{r+1} = 0 \]

which, cannot equal \( 1/p \).

In our case, the success probability (\( p \)) depends on total probability of 1-3 repeats occurring in a defined DNA length, and can be empirically determined from simulated quadruplex formation. As an example, for 120 nt of 3’-ssDNA: total probability = 0.28 (1 repeat) + 0.23 (2 repeats) + 0.14 (3 repeats) = 0.65. From simulation, we can also calculate \( n \) total gaps in \( m \) 3’-ssDNA nucleotides from the average gap size (Ave) expanded to the unit size.

\[
n = \frac{m}{6} * \frac{1}{\text{average unit size for } \frac{m}{6} \text{ repeats}}
\]

where \( m/6 \) is the maximum number of G-repeats in \( m \) nucleotides of 3’-ssDNA. As an example, for 120 nt of 3’-ssDNA containing a hexameric repeat (TTAGGG): \( (120 + 6) / 4 \) (repeats per G-quadruplex) + 1.15 (average repeats between G-quadruplexes) = 3.88. For simplicity, 3.88 is rounded to the nearest integer 4; effectively making the resulting probability (\( P \)) calculation a minimum.

From eq. 1 we can then determine the probability of at least 4 consecutive gaps in \( m \) nucleotides. As an example, for 120 nt: \( q * p^4 = 0.35 * (0.65)^4 = 0.06; x = 1.10 \); leading to a probability \( P(r=4,n=4) = 0.16 \). These results suggest that within a population of 120 nt 3’-ssDNA 16% will stochastically condense into a higher order G-strand structures. As expected, eq. 1 becomes more accurate as \( n \) increase. The probability of higher-order G-strand structure versus 3’-ssDNA length is plotted.

**SUPPORTING INFORMATION**

Supporting Information is available online.
ACKNOWLEDGEMENTS
The authors would like to thank Prof. Ashutosh Tripathi (UNC Macromolecular Interactions Facility) for help with G4-FID studies, Gayan Senavirathne for purified human RAD51 protein, Wendell Jones for G-quadruplex folding simulation and probability guidance and laboratory colleagues for many helpful discussions. This work was supported by grants from the National Institutes of Health to J.D.G (GM31819, ES013773) and R.F (GM080176).

AUTHOR INFORMATION
The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Jack Griffith (jdg@med.unc.edu) or Richard Fishel (fishel.7@osu.edu)

AUTHOR CONTRIBUTIONS
All authors participated in the design of the experiments and the analysis of the data; A.K. isolated the rolling circle DNAs, A.K and N.O.A designed and generated the ssDNA’s; A.K. performed dye exclusion experiments; A.K. and N.O.A. performed the CD analysis; N.D.J. performed the smMT and in silico modeling studies; the EM preparations were done by N.O.A, A.K. and J.D.G; A.K., N.D.J., R.F. and J.D.G wrote the paper and all authors participated in critical discussions.

REFERENCES
1. Fujioka, K., Aratani, Y., Kusano, K., and Koyama, H. (1993) Targeted recombination with single-stranded DNA vectors in mammalian cells. Nucleic Acids Research 21, 407-412
2. Mitra, S., and Kornberg, A. (1966) Enzymatic Mechanisms of DNA Replication. The Journal of General Physiology 49, 59-79
3. Tomonaga, T., and Levens, D. (1996) Activating transcription from single stranded DNA. Proceedings of the National Academy of Sciences of the United States of America 93, 5830-5835
4. Ma, W., Westmoreland, J. W., and Resnick, M. A. (2013) Homologous recombination rescues ssDNA gaps generated by nucleotide excision repair and reduced translesion DNA synthesis in yeast G2 cells. Proc Natl Acad Sci U S A 110, E2895-2904
5. Goddard, N. L., Bonnet, G., Krichevsky, O., and Libchaber, A. (2000) Sequence dependent rigidity of single stranded DNA. Phys Rev Lett 85, 2400-2403
6. Zhang, Y., Zhou, H., and Ou-Yang, Z. C. (2001) Stretching single-stranded DNA: interplay of electrostatic, base-pairing, and base-pair stacking interactions. Biophys J 81, 1133-1143
7. Blackburn, E. H. (1991) Structure and function of telomeres. Nature 350, 569-573
8. Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature 334, 364-366
9. Nicholas, J., Zong, J. C., Alcendor, D. J., Ciufo, D. M., Poole, L. J., Sarisky, R. T., Chiou, C. J., Zhang, X., Wan, X., Guo, H. G., Reitz, M. S., and Hayward, G. S. (1998) Novel organizational features, captured cellular genes, and strain variability within the genome of KSHV/HHV8. J Natl Cancer Inst Monogr, 79-88
10. Rademakers, R., Neumann, M., and Mackenzie, I. R. (2012) Advances in understanding the molecular basis of frontotemporal dementia. Nat Rev Neurol 8, 423-434
11. 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
12. Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012) DNA secondary structures: stability and function of G-
quadruplex structures. Nat Rev Genet 13, 770-780
13. Rhodes, D., and Lipps, H. J. (2015) G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res 43, 8627-8637
14. Sen, D., and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. Nature 344, 410-414
15. Sundquist, W. I., and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature 342, 825-829
16. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell 59, 871-880
17. Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D., and Lipps, H. J. (2008) Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. Nat Struct Mol Biol 15, 598-604
18. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proc Natl Acad Sci U S A 98, 8572-8577
19. Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem 5, 182-186
20. Palm, W., and de Lange, T. (2008) How shelterin protects mammalian telomeres. Annu Rev Genet 42, 301-334
21. Parkinson, G. N., Lee, M. P., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417, 876-880
22. Li, J., Correia, J. J., Wang, L., Trent, J. O., and Chaires, J. B. (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. Nucleic Acids Res 33, 4649-4659
23. Petracco, L. (2013) Higher-order quadruplex structures. Topics in current chemistry 330, 23-46
24. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318, 798-801
25. Schoeftner, S., and Blasco, M. A. (2008) Developmentally regulated transcription of mammalian telomerases by DNA-dependent RNA polymerase II. Nat Cell Biol 10, 228-236
26. Randall, A., and Griffith, J. D. (2009) Structure of long telomeric RNA transcripts: the G-rich RNA forms a compact repeating structure containing G-quartets. J Biol Chem 284, 13980-13986
27. Rippe, K., and Luke, B. (2015) TERRA and the state of the telomere. Nat Struct Mol Biol 22, 853-858
28. Kar, A., Willcox, S., and Griffith, J. D. (2016) Transcription of telomeric DNA leads to high levels of homologous recombination and t-loops. Nucleic Acids Res 44, 9369-9380
29. Yu, H., Gu, X., Nakano, S., Miyoshi, D., and Sugimoto, N. (2012) Beads-on-a-string structure of long telomeric DNAs under molecular crowding conditions. J Am Chem Soc 134, 20060-20069
30. Paramasivan, S., Rujan, I., and Bolton, P. H. (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. Methods 43, 324-331
31. Monchaud, D., Allain, C., and Teulade-Fichou, M. P. (2006) Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. Bioorg Med Chem Lett 16, 4842-4845
32. Monchaud, D., and Teulade-Fichou, M. P. (2010) G4-FID: a fluorescent DNA probe displacement assay for rapid evaluation of quadruplex ligands.
33. Largy, E., Saettel, N., Hamon, F., Dubruille, S., and Teulade-Fichou, M. P. (2012) Screening of a chemical library by HT-G4-FID for discovery of selective G-quadruplex binders. *Current pharmaceutical design* **18**, 1992-2001

34. Burger, A. M., Dai, F., Schultes, C. M., Reszka, A. P., Moore, M. J., Double, J. A., and Neidle, S. (2005) The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer research* **65**, 1489-1496

35. Mo, Y., Gan, Y., Song, S., Johnston, J., Xiao, X., Wientjes, M. G., and Au, J. L. (2003) Simultaneous targeting of telomeres and telomerase as a cancer therapeutic approach. *Cancer research* **63**, 579-585

36. Campbell, N. H., Parkinson, G. N., Reszka, A. P., and Neidle, S. (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *J Am Chem Soc* **130**, 6722-6724

37. Parkinson, G. N., Ghosh, R., and Neidle, S. (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry* **46**, 2390-2397

38. Chi, Q. J., Wang, G. X., and Jiang, J. H. (2013) The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Physica A* **392**, 1072-1079

39. Gosse, C., and Croquette, V. (2002) Magnetic Tweezers: micromanipulation and force measurement at the molecular level. *Biophys J* **82**, 3314-3329

40. Lipfert, J., Hao, X., and Dekker, N. H. (2009) Quantitative modeling and optimization of magnetic tweezers. *Biophys J* **96**, 5040-5049

41. De Vlaminck, I., Henighan, T., van Loenhout, M. T., Burnham, D. R., and Dekker, C. (2012) Magnetic forces and DNA mechanics in multiplexed magnetic tweezers. *PLoS One* **7**, e41432

42. Lionnet, T., Allemand, J. F., Revyakin, A., Strick, T. R., Saleh, O. A., Bensimon, D., and Croquette, V. (2012) Magnetic trap construction. *Cold Spring Harb Protoc* **2012**, 133-138

43. Savitzky, A., and Golay, M. J. E. (1964) Smoothing + Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **36**, 1627-&

44. Reineker, P., Winkler, R. G., and Glattling, G. (1995) Freely jointed chain with variable segment number and length. *Colloid and Polymer Science* **273**, 32-37

45. Bouchiat, C., Wang, M. D., Allemand, J., Strick, T., Block, S. M., and Croquette, V. (1999) Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys J* **76**, 409-413

46. Freedman, D. A., Pisiani, R., and Purves, R. A. (2007) *Statistics*, 4th Edition ed., W.W. Norton, Inc., New York, N.Y.

47. de Messieres, M., Chang, J. C., Brawn-Cinani, B., and La Porta, A. (2012) Single-molecule study of G-quadruplex disruption using dynamic force spectroscopy. *Phys Rev Lett* **109**, 058101

48. Lynch, S., Baker, H., Byker, S. G., Zhou, D., and Sinniah, K. (2009) Single molecule force spectroscopy on G-quadruplex DNA. *Chemistry* **15**, 8113-8116

49. 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

50. Wang, Q., Ma, L., Hao, Y.-h., and Tan, Z. (2010) Folding Equilibrium Constants of Telomere G-Quadruplexes in Free State or Associated with Proteins Determined by Isothermal Differential
Hybridization. *Analytical Chemistry* **82**, 9469-9475

51. Hänsel, R., Löhr, F., Foldynová-Trantírková, S., Bamberg, E., Trantírek, L., and Dötsch, V. (2011) The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Research* **39**, 5768-5775

52. Qureshi, M. H., Ray, S., Sewell, A. L., Basu, S., and Balci, H. (2012) Replication Protein A Unfolds G-Quadruplex Structures with Varying Degrees of Efficiency. *J Phys Chem B* **116**, 5588-5594

53. Tombline, G., Heinen Cd Fau - Shim, K.-S., Shim Ks Fau - Fishel, R., and Fishel, R. Biochemical characterization of the human RAD51 protein. III. Modulation of DNA binding by adenosine nucleotides.

54. Heyer, W. D., Ehmse Kt Fau - Liu, J., and Liu, J. (2010) Regulation of homologous recombination in eukaryotes.

55. Cesare, A. J., and Reddel, R. R. (2010) Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet* **11**, 319-330

56. Senaviratne, G., Mahto, S. K., Hanne, J., O'Brian, D., and Fishel, R. (2017) Dynamic unwrapping of nucleosomes by HsRAD51 that includes sliding and rotational motion of histone octamers. *Nucleic Acids Research* **45**, 685-698

57. Conway, A. B., Lynch, T. W., Zhang, Y., Fortin, G. S., Fung, C. W., Symington, L. S., and Rice, P. A. (2004) Crystal structure of a Rad51 filament. *Nat Struct Mol Biol* **11**, 791-796

58. Shinohara, A., and Ogawa, T. (1999) Rad51/RecA protein families and the associated proteins in eukaryotes. *Mutation research* **435**, 13-21

59. Brockman, C., Kim, S. J., and Schroeder, C. M. (2011) Direct observation of single flexible polymers using single stranded DNA. *Soft Matter* **7**, 8005-8012

60. Lee, K. S., Marcil, A. B., Kozlov, A. G., Schroeder, C. M., Lohman, T. M., and Ha, T. (2014) Ultrafast Redistribution of E. coli SSB along Long Single-Stranded DNA via Intersegment Transfer. *Journal of molecular biology* **426**, 2413-2421

61. Marsh, T. C., and Henderson, E. (1994) G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTGGGG), into large superstructures. *Biochemistry* **33**, 10718-10724

62. Feller, W. (1968) An Introduction to Probability Theory and Its Applications. *Wiley* 1

63. 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

64. Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D., and Shay, J. W. (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev* **11**, 2801-2809

65. Arnoult, N., and Karlseder, J. (2015) Complex interactions between the DNA-damage response and mammalian telomeres. *Nat Struct Mol Biol* **22**, 859-866

66. Taylor, D. J., Podell, E. R., Taatjes, D. J., and Cech, T. R. (2011) Multiple POT1-TPP1 proteins coat and compact long telomeric single-stranded DNA. *Journal of molecular biology* **410**, 10-17

67. Thys, R. G., and Wang, Y. H. (2015) DNA Replication Dynamics of the GGGGCC Repeat of the C9orf72 Gene. *J Biol Chem* **290**, 28953-28962

68. Griffith, J. D., and Christiansen, G. (1978) Electron microscope visualization of chromatin and other DNA-protein complexes. *Annu Rev Biophys Bioeng* **7**, 19-35
**FIGURE LEGENDS**

**Figure 1. Fluorescence Intercalator Displacement Analysis of G-quadruplex DNA.** a) Schematic representation of G-quadruplex fluorescence intercalator displacement (G4-FID) analysis. The displacement of thiazole orange (TO) by the G-quadruplex binding-specific BRACO-19 ligand results in reduced fluorescence intensity. b) G4-FID analysis performed with 24 nt (G24), 120 nt (G120) and rolling circle-generated (G-strand) G-strand DNA. c) G4-FID analysis of rolling circle-generated G-strand, C-strand and mutant G-strand DNA.

**Figure 2. Electron Microscopic Visualization of long telomeric ssDNAs.** Long telomeric ssDNA was synthesized by rolling circle replication and prepared for EM by direct adsorption to thin carbon supports, dehydration, and rotary metal shadow casting with tungsten. a) C-strand ssDNA. b) Mutant (TTAGTG)n G-strand ssDNA. c) G-strand DNA. d) Higher magnification of G-strand DNA. Small arrowheads indicate small bead-like structures; large arrowheads indicate large bead-like structures.

**Figure 3. Size and Length Distribution of G-strand Bead-Like Structures.** a) The diameter of G-strand bead-like structures was determined by EM (N = 271; see Figure 2). The binned histogram was fit to two Gaussian curves and the mean and standard deviation (s.d.) determined (see Experimental Procedures). b) Representative EM examples of G-strand DNA saturated with large bead-like particles (left). Bar is equivalent to 50 nm. Box plot of length measures determined by EM for G-strand DNA saturated with large bead-like particles (right; N = 91). c) Binned histogram of individual G-strand DNA saturated with large bead-like particles length measures divided by the number of bead-like particles observed on each individual molecule (N = 91).

**Figure 4. Single Molecule Magnetic Tweezer Force Extension Analysis of G-strand ssDNA.** a) Force extension analysis of rolling circle and size-selected 5 knt human G-strand ssDNA (N = 25, blue squares), size-selected 5 knt mutant G-strand ssDNA (N = 22, black squares), and 5 kb dsDNA (N = 28, red square). Red box indicates forces where discrete steps are observed during force extension. b) Representative extension versus time (frame number at a frame rate of 14.2 msec/frame) of 5 knt G-strand ssDNA (black) and mutant G-strand ssDNA (red). Red box indicates forces/frames shown in red box outlined in Panel a. c) Representative magnification of discrete steps observed in red boxes of Panels a and b. Red triangles indicate the time of magnet steps (2 msec movement). Blue lines indicate large change in the second derivative of movement where extension steps are determined. d) Binned 5 knt G-strand ssDNA extension steps (N = 434; 62 individual DNA molecules). Multiple Gaussian peaks are shown in different colors with the smoothed curve shown in black. Each Gaussian peak is labeled with a letter (a-f) and the mean step size with standard deviation.

**Figure 5. Hypothetical Model of Higher-Order G-Strand Structures.** a) Caricature illustration of G-quartet as a black box, extended connector non-G-residue nucleotides between G-quadruplex folds (red) and non-folded G-rich repeats between G-quadruplex folds (blue dots). G-rich repeats may fold into several types of G-quadruplex structures (parallel, anti-parallel, hybrid, etc.) with non-G-residue nucleotide loops connecting the corners of the hypothetical box. In an array of G-quadruplex folds, regardless of the type of folded structure, the non-G-residue nucleotides of the last G-rich repeat may extend to the adjacent G-quadruplex. For the human telomeric repeat (TTAGGG)n, the non-G-residue nucleotides TTA (red) comprise this extended connection, with on average at least one non-folded G-rich repeat (TTAGGG; blue dots). b) The results suggest that multiples of four non-folded intervening G-rich repeats (blue dots) may condense into G-residue pseudo-folds (blue box). The
condensation into G-residue pseudo-folds is expected to be completely independent of the type of G-quartet fold(s) within the G-quadruplex array since they involve only the intervening G-rich repeats. c) Under magnetic force G-residue pseudo-folds may extend in multiples of ~29 nm.

Figure 6. The binding of RAD51 to G-Strand ssDNA. a) Representative force-extension of a single G-strand ssDNA in which the applied force is released following extension and then re-applied for a total of 3 successive cycles. b) Representative force-extension of a single G-strand ssDNA in which the applied force is released following extension and then re-applied for a total of 3 successive analysis in the presence of human RAD51 and in the absence of ATP (-ATP). c) Representative force-extension of a single G-strand ssDNA in which the applied force is released following extension and then re-applied for a total of 3 successive analysis and in the presence of ATP (+ATP).

Figure 7. Visualization of RAD51 extension of G-strand DNA. Long G-strand ssDNA was incubated with RAD51 protein in the presence of ATP and then prepared for EM by negative staining (Experimental Procedures). The ssDNA appears highly extended in agreement with the smMT analysis and the helical pitch in the filaments (107 A) is typical of RAD51-ssDNA complexes formed with ATP.
Figure 1

a) Diagram showing high fluorescence intensity and low fluorescence intensity.

b) Graph showing TO displacement (%): G24, G120, G-strand.

To: Ligand Concentration (μM)

TO Displacement (%)

0.00 0.25 0.50 0.75 1.00 1.25 1.50

0 20 40 60 80 100

G-strand

C-strand

Mutant G-strand

Graph showing TO displacement (%): G-strand, C-strand, Mutant G-strand.

To: Ligand Concentration (μM)

TO Displacement (%)

0.00 0.25 0.50 0.75 1.00 1.25 1.50

0 20 40 60 80 100
Figure 3

(a) Bar graph showing the size distribution with a peak at 6.1 nm ± 1.5 s.d. and another peak at 9.4 nm ± 1.9 s.d.

(b) Image showing two samples with size data.

(c) Bar graph showing the length distribution with a peak at 13.7 nm/bead ± 1.9 s.d.
Figure 4

(a) Force (pN) vs. Extension (nm) for 5 knt G-strand and 5 knt Mutant G-strand. The inset highlights the step size distribution for these strands.

(b) Extension (nm) vs. Time (Frame Number) for 5 knt G-strand and 5 knt Mutant G-strand. The inset is a close-up of the data points.

(c) Extension (nm) vs. Time (Frame Number) showing 5 kb dsDNA.

(d) Step Size distribution for 62 DNA molecules. The step sizes are labeled as follows: a = 33 nm ± 14, b = 55 nm ± 14, c = 77 nm ± 14, d = 99 nm ± 14, e = 117 nm ± 14, f = 142 nm ± 14.
Figure 5

(a) (G G G T T A)_{0-3}

(b) Force-Extension

(c) (4.2 nm)_{0-3}

~29 nm

2.1 nm 0.6-1.5 nm
Figure 6

(a) G-strand ssDNA only

(b) RAD51 (-ATP)

(c) RAD51 (+ATP)
Long repeating (TTAGGG)n single stranded DNA self-condenses into compact beaded filaments stabilized by G-quadruplex formation
Anirban Kar, Nathan Jones, N Özlem Arat, Richard Fishel and Jack Griffith

*J. Biol. Chem.* published online April 19, 2018

Access the most updated version of this article at doi: 10.1074/jbc.RA118.002158

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts