Identification of the New Type of G-Quadruplex with Multiple Vacant Sites in Human Telomeric DNA

Xiaonong Zhang†, Xiaolong Kou†, Wei Zhang & Wenke Zhang*

State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry, Jilin University, Changchun 130012

*Corresponding author: zhangwk@jlu.edu.cn; †X. Zhang and X. Kou contributed equally to this work.

Cite this: CCS Chem. 2021, 3, 3192–3204

To date, few studies have reported on the folding mechanism of tandem G-quadruplexes in human telomeric DNA. Hence, the control of the biofunctions of G-quadruplex, which requires a thorough understanding of its dynamic behavior, is limited. Here, we investigated the folding/unfolding behavior of human telomeric sequences with lengths over 10 kilonucleotide (knt) by circular dichroism (CD) spectroscopy, UV melting assay, and atomic force microscopy (AFM)-based single-molecule force spectroscopy. A novel G-quadruplex with multiple vacant sites was captured in the long human telomeric DNA and denoted as pre-G-quadruplex (pre-GQ). According to the number of vacant sites, pre-GQ is divided into two types (four vacant sites in type 1 and two vacant sites in type 2), among which type 1 is the dominant one. The unfolding force of a tandem pre-GQ was 10 pN lower than that of a complete G-quadruplex, suggesting the destabilized structure of the tandem pre-GQ due to its incompletely folded state. Our results revealed that the folding of long-telomere G-strand could be achieved by a two-step process involving fast transition (in seconds) from unstructured ssDNA to tandem G-hairpin states and slow folding of free guanines into the vacant sites in tetraplex to generate type 1 pre-GQ, type 2 pre-GQ, or complete GQ.

Keywords: telomere G-quadruplex, long telomere DNA, single-molecule force spectroscopy, vacant site, folding mechanism

Introduction

G-quadruplex, a four-stranded structure generated by Hoogsteen hydrogen bonding in G-rich telomere and some promoter regions,1−5 has been found to participate in tumor cell apoptosis via telomerase activity inhibition.6,7 G-quadruplex has attracted considerable attention as a potential target for cancer therapy,8−11 and numerous investigations have been carried out to acquire accurate information regarding its structure and folding process. Parallel (four G-rich strands arranged in the same direction), antiparallel (two strands folded in the same direction), and parallel/antiparallel hybrid (one strand assembled in the opposite direction to the other three strands) conformations were observed.12−15 The single G-quadruplex tends to adopt an antiparallel conformation and a hybrid parallel/antiparallel conformation in Na+ and K+ buffers, respectively.12,15 Both reduction of K+ concentration and temperature elevation can lead to unfolding of G-quadruplex, which experiences a triplex conformation...
intermediate driven by positive heat capacity change.16–20 Human telomeric DNA and its variants have attracted the attention of many researchers, and a variety of the folding modes/processes have been identified.21–24

Human telomeric DNA is composed of thousands of (TTAGGG)n/(CCCTAA)n repeats,25 and the G-rich strand forms a telomere overhang that typically ranges from 100 to 300 nucleotides,26 which enables in vivo formation of tandem G-quadruplexes. To explore the practical structure of human telomere overhang, tandem G-quadruplexes with 2 to 5 G-quadruplex forming sequences were prepared and investigated using circular dichroism (CD) spectroscopy. Some studies have suggested that G-quadruplexes can move independently in the multiquadruplex structures as a consequence of limited intermolecular interactions. This has been described as “beads-on-a-string.”27,28 However, stacking interactions between G-tetrads cores or the TTA loop have been reported in recent studies.29–33 These controversial findings may be attributed to the differences in the buffer compositions and the lengths of the G-rich sequences.34

Single-molecule force spectroscopy, a promising technique allowing constant and dynamic probing of minute forces and lengths at the single-molecule level with piconewton and subnanometer resolution,35–41 has been employed to explore the intramolecular and intermolecular interactions of G-quadruplexes.19,30,42–46 Using 144 nt human telomeric G-rich sequences, Mao et al.43 reported that quadruplex-quadruplex interaction (QQI) was detected in only 5% of all G-quadruplexes, while the majority of G-quadruplexes followed the beads-on-a-string model. However, with an ~5000 nt telomeric G-rich strand as the research object, Griffith et al.47 reported interactions of adjacent G-quadruplexes, which induced self-condensation of G-quadruplexes into compact beaded filaments. Additionally, unfolding of partially folded structures has been observed in these studies, and they are classified as G-hairpin or G-triplex.43,48 Besides these traditional partially folded structures, G-quadruplexes with (4n-1) guanines in the G-tetrad core have also recently been reported.49 In summary, despite considerable attempts in understanding the folding/unfolding mechanism of tandem G-quadruplexes, the practical structure and dynamic folding/unfolding process of full-length human telomere overhang are still uncertain.

In this study, UV melting assay, CD spectroscopy, and single-molecule force spectroscopy (SMFS) were employed to investigate the conformation and the folding/unfolding process of telomeric long G-rich sequences prepared by rolling circle amplification (RCA).50 A novel partially folded G-quadruplex, which displays similar CD signals with those of G-hairpin and 10 pN reduction in unfolding force compared to a single complete G-quadruplex, was identified in the long human telomeric sequence. This partially folded G-quadruplex, which is distinct from G-intermediates reported elsewhere,51,52 is denoted as pre-G-quadruplex (pre-GQ). Pre-GQ is categorized according to the number of its vacant sites and has two possible structures: type 1 (four vacant sites) and type 2 (two vacant sites). The long-telomeric sequences are rapidly assembled into tandem G-hairpins and then slowly folded into type 1 pre-GQs, type 2 pre-GQs, or complete GQs, which require an incubation time of at least 10 min. The folding process can be accelerated by the presence of polyethylene glycol (PEG). This study facilitates the understanding of the structure and dynamic folding mechanism of long-chain tandem G-quadruplexes.

**Experimental Methods**

**Materials and reagents**

The high-performance liquid chromatography (HPLC)-purified DNA oligomers were purchased from Sangon Biotech Co., Ltd. (Shanghai, China). All sequences used in this study are summarized in Supporting Information Tables S1–S3. A CircLigase ssDNA Ligase (CL 4115K) kit was purchased from Epicenter Biotechnologies (Madison, WI). φ29 DNA polymerase required for RCA was purchased from New England Biolabs ( Ipswich, MA). Tris(2-carboxyethyl)phosphine (TCEP), 6-Mercaptohexan-1-ol (MCH), and PEG 200 were purchased from Sigma-Aldrich Co. LLC (St. Louis, MO, USA). All chemical reagents were of analytical grade and used without further purification.

**RCA**

The long-telomeric ssDNA used in CD spectroscopy, UV melting, and atomic force microscopy (AFM) imaging were prepared by RCA through the procedures reported previously.50 The circle templates were synthesized by ligating 25 μM of 5′-phosphorylated oligonucleotides (liner templates, T1 or T2) in 40 μL reaction buffer containing 50 mM 3-(N-morpholino)propanesulfonic acid (MOPS) (pH 7.5), 10 mM KCl, 50 mM MgCl2, 10 mM dithiothreitol (DTT), 100 μM adenosine triphosphate (ATP), 2.5 mM MnCl2, and 200 U CircLigase ssDNA ligase. To achieve the cyclization reaction, the mixture obtained was cultured at 60 °C for 10 h, followed by treatment at 80 °C for 10 min to denature the enzymes. For the RCA reaction, 1.8 μM of circle DNA templates and 0.7 μM of ssDNA primers (P1 or P2) in 455 μL reaction buffer (containing 50 mM Tris–HCl, 10 mM MgCl2, 10 mM (NH4)2SO4, 4 mM DTT, pH 7.5) were annealed at 95 °C for 3 min and then kept at 0 °C for 15 min. Then, 150 U of φ29 DNA polymerase and 0.2 mmol deoxy-ribonucleoside triphosphate (dNTPs) were added into the mixture, and the final volume of the mixture was 490 μL. The
mixture was incubated at 30 °C overnight, followed by incubation at 65 °C for 10 min. Finally, the RCA products were purified by alcohol precipitation and stored at −20 °C.

**CD measurement**

For CD measurement, ssDNA oligonucleotides were diluted with 20 mM Tris–HCl (pH 7) containing 100 mM NaCl or 100 mM KCl to achieve a concentration of 5 μM for short ssDNA and ~200 μg/mL for long ssDNA (RCA products). The sample was transferred into a 0.5 mm path length quartz cell, followed by CD measurements using a MOS-450 spectrometer (Bio-logic, France). All data were collected from 220 to 340 nm with a scanning rate of 0.5 nm/s and analyzed by Igor Pro (WaveMetrics, Portland, USA) using a Savitzky-Golay function.

**UV-melting measurement**

ssDNA samples were diluted with 20 mM Tris–HCl (pH 7) containing 100 mM KCl to achieve a concentration of 5 μM for short ssDNA and ~200 μg/mL for long ssDNA (RCA products). The sample was placed into a 0.5 cm path length quartz cell and loaded onto a UV 2450 spectrometer (Shimadzu, Japan). The measurements were executed at 295 nm, with a temperature range of 10–95 °C and a heating rate of 0.2 °C/min. All data were analyzed by Igor Pro (WaveMetrics, Portland, USA).

**AFM imaging**

DNA samples (nhtel and interval-nhtel) were diluted with a buffer containing 10 mM MgCl$_2$, 100 mM KCl, and 10 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) (pH 7). The samples were then incubated at 90 °C for 10 min (for denaturation) and at 25 °C for 20 h (for the formation of G-quadruplexes). Before AFM imaging, the sample was deposited onto freshly cleaved mica, followed by rinsing with milli-Q water and drying by nitrogen gas. AFM imaging was achieved using NanoWizardII BioAFM (JPK Instruments AG (now Bruker), Berlin, Germany) with silicon cantilevers (OTESPA-R3, 26 N/m, Bruker, Germany). All images were collected in air and analyzed by JPK imaging software.

**Preparation of ssDNA-modified substrate**

The Au-coated glass slide was cut into 1.2 × 1.2 cm$^2$ pieces and treated with freshly prepared piranha solution (H$_2$SO$_4$/H$_2$O$_2$ = 7.3, v/v) for 30 min. The slides were then rinsed thoroughly with deionized water and dried by nitrogen gas.

The ssDNA primers (sequences of all primers are shown in Supporting Information Table S1) were coupled to the Au-coated substrate via thiol-gold chemistry.$^{52}$ In brief, a drop of thiol-labeled ssDNA primer solution containing 1 nM of 5′-thiol-labeled primer, 0.2 nM of 3′-thiol-labeled spacer DNA T20 oligo, and 0.7 mM of TCEP was placed on the Au-coated substrate and incubated at room temperature for 12 h. The substrate was then rinsed thoroughly with deionized water.

To obtain long ssDNA-modified Au-coated substrate, the RCA reaction was performed on the primer-modified Au-coated substrate. Specifically, the circle templates (0.25 nM) were added into a 1 × Φ29 reaction buffer (50 mM of Tris–HCl, 10 mM of MgCl$_2$, 40 mM of (NH$_4$)$_2$SO$_4$, 4 mM of DTT, pH 7.5) and incubated at 95 and 0 °C for 5 min each. Then, 25 U of Φ29 DNA polymerase and 50 pmol of dNTPs were subsequently added to the circle template containing the solution to obtain a reaction mixture with a volume of 70 μL. The mixture was deposited on the primer-modified Au-coated substrate and incubated at 30 °C for 12 h for RCA. The ssDNA-modified substrate was rinsed by hot water (90 °C) to deactivate Φ29 DNA polymerases. 100 μL MCH (0.25 M) was dropped on the substrate to destroy the physical adsorption of DNA molecules on the substrate. After incubation for 15–20 min, the sample was rinsed by ultrapure water and used immediately for force spectroscopy assay.

**Characterization by AFM-based single-molecule force spectroscopy**

The force spectroscopy investigations were performed using a ForceRobot 300 (JPK Instruments AG, Berlin, Germany) in contact mode, as described previously.$^{53}$

The spring constants of cantilevers (BL-RC150B-C1, Olympus, Japan), which ranged from 26 to 38 pN/nm, were calibrated by the thermal noise method.$^{54}$

For the single-time stretching experiment, the tip was pressed at the substrate with a contact force of 0.5 nN and contact time of 2 s to allow sample adsorption on the tip. The sample was then stretched at a rate of 1 μm/s. Before the first stretching curve, system stabilization, parameter setting, and spring constant calibration typically take 10 min. Therefore, long-chain nhtel has around 10 min to fold into high-order structures in the buffer.

For the stretching-relaxation experiment, the tip was pressed at the substrate with a contact force of 1–2 nN and contact time of 30 s. Then, the molecules were stretched and relaxed at different rates (0.2, 0.5, 1, 2, and 5 μm/s) for at least 40 cycles. The stretching length gradually increased from 1 to 5 μm. The interval between each stretching/relaxation cycle ranged from 0 to 60 s. Data analysis, including curvilinear integral and worm-like chain (WLC) fitting,$^{55}$ was performed using self-written macros embedded in Igor Pro (WaveMetrics). The histograms of plateau force were fitted by the Gaussian function to obtain the accurate results.
Results and Discussion

G-vacant may exist in the structures formed by long-telomeric G-strand

Long-chain telomeric G-strand ssDNA (nhtel) was synthesized by RCA (see details in the “Experimental Methods” section and Supporting Information Figure S1). The long-telomeric ssDNA was cultured in 100 mM KCl buffer at 25 °C for 20 h after thermal denaturation to allow formation of tandem G-quadruplexes. As shown in Figure 1a (white arrowheads) and Supporting Information Figure S2, beads-on-a-string structures were observed for nhtel. These structures exhibited an average height of 1.4 ± 0.2 nm (n = 231). The black curve indicates the Gaussian fit of the height.

Figure 1 | Visualization of nhtel by AFM. Long-telomeric G-strand ssDNA (nhtel) was synthesized by RCA. (a) AFM imaging of nhtel after incubation in 100 mM KCl buffer (pH 7). White arrowheads indicate small bead-like structures. (b) Height distribution of G-strand bead-like structures is centered at 1.4 ± 0.2 nm (n = 231). The black curve indicates the Gaussian fit of the height.

CD spectroscopy was employed to investigate the structure of the long-telomeric G-rich sequence. Herein, short telomeric G-rich sequences allowing the formation of one (1htel), two (2htel), three (3htel), and five (5htel) G-quadruplexes were used as the control group. The CD curves in Figure 2a were obtained after incubation of G-rich sequences (1htel and nhtel) in 100 mM KCl buffer at 25 °C for 20 h. 1htel exhibited a typical CD signal of hybrid G-quadruplex (yellow, Figure 2a), which possessed positive bands at 245, 265, and 295 nm. Additionnally, a positive peak at 270 nm is present in the CD spectrum of nhtel (red, Figure 2a).

The position of the 270-nm peak was independent of the buffer composition (Supporting Information Figure S3). This was different from reported tetraplex (parallel, antiparallel, or hybrid), but similar to G-hairpin, which exhibited a positive band at 270 nm and a negative band at 245 nm. The stable G-hairpin structure in the long-telomeric chain was a surprising finding. As a suggested intermediate in G-quadruplex folding, G-hairpin is highly unstable and difficult to detect. In addition, UV melting analysis showed a reduced thermal stability for nhtel (Figure 2b). These results suggested the presence of a unique structure in the long-telomeric sequence, which may correspond to a novel partially folded G-quadruplex.

To investigate the effect of chain length on the folding behavior of telomeric ssDNA, we compared the structures formed by G-rich sequences under two extreme conditions: (1) at 0 °C for 5 min and (2) at 25 °C for 20 h. Figures 3a and 3b show the CD curves of 1htel, 2htel, 3htel, and 5htel in 100 mM KCl buffer. All CD curves in Figure 3a were obtained by incubating G-rich sequences in 100 mM KCl buffer at 0 °C for 5 min and measured immediately, leaving no sufficient time for folding of G-rich sequences. Indeed, only 1htel can form hybrid tetraplex in this case. Different from 1htel, which is the shortest chain, 2htel, 3htel, and 5htel exhibited a significant negative peak at 260 nm, besides a positive peak at 290 nm, which was similar to the typical CD

Figure 2 | The comparisons on the CD and UV spectra of long G-strand (nhtel) and single G-quadruplex (1htel). (a) CD spectra of 1htel (yellow) and nhtel (red) in solution with 100 mM of KCl and pH 7. (b) UV melting curves of 1htel (yellow) and nhtel (red) in solution with 100 mM of KCl. Before the measurements, all samples were incubated at 25 °C for 20 h after thermal denaturation.
However, after folding at higher temperature (25 °C) for a much longer time (20 h), 2htel, 3htel, and 5htel can form hybrid tetraplex (see Figure 3b and Supporting Information Figure S4). Therefore, we propose that long-telomeric G-strands need more time to fold into stable structures, and the final folding structures of long G-stands \((n \geq 2, \text{in } nhtel)\) will have partially folded structures. To verify this hypothesis, we designed mutated sequences M01–M10 and M11 via mutation of two guanines (Gs) to thymines (Ts). The G-quadruplex formed by these two G-rich fragments should exhibit vacant sites (grey rectangles, Figure 3d). As expected, the CD spectra of long-telomeric ssDNA may have unfolded guanines at the 3′ or 5′ end, resulting in the formation of tetraplex with vacant sites. This may be an important reason for the low thermodynamic stability and the unique CD signal of nhtel in 100 mM KCl buffer, as shown in Figure 2.

**Pre-G-quadruplex as a novel G-quadruplex with multiple vacant sites**

The CD spectra of nhtel displayed a positive broad band at 270 nm in both KCl and NaCl buffers (Supporting Information Figure S3). As discussed above, long-telomeric G-strand has difficulty in forming complete tetraplexes and tends to form tetraplex with vacant sites. With G-quadruplex-binding ligands (TMPyP4) added,\(^5^9\) positive bands at 265 and 290 nm can be observed (Supporting Information Figure S5), demonstrating that nhtel can form tandem partially folded G-quadruplexes in KCl buffer, while the unfolded guanines fold to vacant sites with the help of GQ ligands. Compared to mutated sequences M11 and M01–M10 (Figures 3c and 3d), the structure formed by nhtel may have more than one free guanine at the 3′ or 5′ end.

Mutated sequences (Mxy) were designed to verify our hypothesis (Supporting Information Table S3). Herein, \(x\) and \(y\) represent the number of G-to-T mutations on the 5′ and 3′ ends of 1htel, respectively. As shown in Supporting Information Figures S6 and S7, mutated sequences M01, M11, and M10 \((x \leq 1 \text{ and } y \leq 1)\) form intramolecular antiparallel G-quadruplex, while mutated sequences M12, M21, M22, M20, and M02 \((x \geq 2 \text{ or } y \geq 2)\) form...

---

**Figure 3** | G-vacant may exist in the structures formed by long G-strand. After thermal denaturation, 1htel (yellow), 2htel (light green), 3htel (blue), and 5htel (purple) were treated by different methods before CD measurements under different conditions. (a) The samples were cooled down in ice water for 5 min and measured immediately. (b) The samples were incubated at 25 °C for 20 h before measurements. (c) CD spectra of mutated sequences M11 (orange) and M01–M10 (dark green). All CD measurements were performed in solutions with 100 mM of KCl and pH 7. (d) The schematic of the structures formed by 2htel at 0 °C and mutated sequences M01–M10 and M11 folded at 25 °C.
intermolecular G-quadruplex. Short mutated G-rich sequences can form intermolecular G-quadruplex due to multiple G-to-T mutations, thus cannot simulate partially folded G-quadruplex formed by long-chain nhtel. Therefore, long mutated G-strands shown in Supporting Information Table S3 were designed. Mutated sequences can be divided as symmetric (5xM11, 5xM22, and 5xM33) and asymmetric (5xM02, 5xM20, 5xM12, 5xM21, and 5xM03) mutations.

As shown in Figure 4, a positive band at 280 nm, which corresponds to the unstructured ssDNA, is present in the CD spectra of symmetrically mutated sequences (5xM22 and 5xM33) in 100 mM KCl buffer (dashed). To facilitate folding of long-telomeric G-strands, a 40% PEG solution (w/w) was added into the system. As the incubation progressed, the positive band at 280 nm blue-shifted to 270 nm, indicating a transition from unstructured ssDNA to G-hairpin structure (Figures 4b and 4c). However, two positive bands at 270 and 290 nm in the CD spectra of asymmetrically mutated sequences (5xM12, 5xM21, 5xM20, and 5xM02) in 100 mM KCl buffer (Supporting Information Figure S8). In the presence of PEG, the intensity of the peak at 270 nm increased, while the intensity of the peak at 290 nm decreased as the incubation progressed. Indeed, the CD spectra of 5xM03 (five tandem G-triplexes) (Supporting Information Figure S8c) in the presence of PEG were highly similar to those of asymmetrical sequences in the absence of PEG. Hence, it is hypothesized that the asymmetric sequences can fold into tandem G-triplexes or partial G-triplexes in the absence of PEG. In the presence of PEG, the extra guanine (red “*” in Supporting Information Figure S9) of the triplex formed by M12, M21, M20, or M02 participated in folding, resulting in the shifts of CD peaks.

After a long incubation (e.g., 24 h) in the presence of PEG, the intensity of the positive peak at 270 nm decreased (as shown by the light blue and solid dark blue line in Figure 4a), and a new positive shoulder-like peak corresponding to the formation of antiparallel quadruplex was observed at 290 nm in CD spectra of nhtel. The CD spectrum of 5xM22 in the presence of PEG after 24-h incubation was highly similar to that of nhtel in the KCl buffer (Figures 4a and 4b), except that the band at 290 nm was not observed, even after incubation for 3 days (data not shown). The results indicated that the antiparallel quadruplex was formed by folding of the unfolded guanines at the 3′ and 5′ ends to the vacant sites, rather than the dimerization of adjacent G-hairpin structures. Compared with asymmetrically mutated sequences, the CD signals of nhtel were more similar to those of the symmetrically mutated sequences. Therefore, the vacant sites on the 5′ and 3′ ends of the partially G-quadruplex formed by long-telomeric G-strands are supposed to be symmetrical. This may be attributed to the fact that the long G-strands

Figure 4 | CD characteristics and possible structure of pre-GQ. CD spectra of (a) nhtel, (b) 5xM22, and (c) 5xM33. The black dashed curves correspond to the CD spectra of samples incubated at 25 °C for 24 h in the absence of PEG after thermal denaturation. The solid curves represent CD spectra of samples incubated at 25 °C for 4 h (light blue) and 24 h (dark blue) in presence of PEG, respectively. All CD measurements were performed in 100 mM KCl buffer with pH 7 at 25 °C. (d) Schematic of possible conformations formed by nhtel. Red rectangle refers to free guanine, blue rectangle refers to folded guanine, green rectangle refers to thymine, and grey rectangle refers to the vacant site.
restrict folding of single G-quadruplex structure, and the restrictions on both sides are approximately equivalent. Additionally, the CD data suggest that long-telomeric G-strands (nhtel) were present in 100 mM of KCl buffer mainly as G-hairpin or G-quadruplex with four vacant sites (type 1) (Figure 4d). In the presence of PEG, some free guanines (G1, G2, G3, G19, G20, G21, shown in Figure 4d) folded into the vacant sites in the partially folded structure, resulting in stable conformations such as G-quadruplex with two vacant sites (type 2) and complete antiparallel tetraplex (Figure 4d). As a result, a shoulder-like peak at 290 nm was observed. However, the peak at 290 nm was relatively weak compared with the one at ~270 nm, suggesting that the majority of telomeric DNA was present as G-hairpin or type 1 structure with a small proportion of type 2 structure, even in such a crowded environment. These novel, partially folded structures are named as pre-GQ. Both pre-GQs exhibit bead-like structures. Since G-hairpin and type 1 pre-GQ have identical characteristic CD signals (a positive band at 270 nm), it is difficult to identify the dominant structure by the CD method. To further explore the pre-GQ structure, we investigated the unfolding/refolding behavior of the structure by using AFM-based SMFS.

The mechanical stability of pre-G-quadruplex

Unfolding of tandem pre-GQs was achieved by AFM-based single-molecule force spectroscopy (Figure 5a). Herein, the thiol-labeled long-telomeric chains prepared as described above were immobilized on the Au-coated substrate and AFM tip by Au–S bonds and physical adsorption, respectively. Figure 5 shows the typical force-extension curves and corresponding histograms of the unfolding forces. The majority (67%) of the force-extension curves obtained were identified by a flat-force plateau with a statistical unfolding force of 40 ± 8 pN (type I, red trace, and histogram in Figures 5b and 5c), which is 10 pN lower than that of a single G-quadruplex. This further supports the proposed hypothesis on the partially folding mode of the novel structure (pre-GQ), and is consistent with the decreased melting temperature in the UV-melting assay.

Besides the flat-force plateau, a tilted plateau with an average force of 59 ± 19 pN (Figure 5c, type II) was also observed during manipulation of the tandem repetitive nhtel chains. The disassembly of high-order structures in the long-telomeric sequences might contribute to the tilted plateau with increased force. To verify this hypothesis, a telomeric DNA with an interval sequence at the 5’ end of each 4G-tracts unit was synthesized by RCA (interval-nhtel). CD curves of the interval-nhtel in KCl buffer showed a positive broad peak at 260–285 nm. This broad peak was divided into two peaks located at ~275 and 285 nm, respectively (Supporting Information Figure S10). The G-hairpin and type 1 pre-GQ formed in nhtel were characterized by a peak at 270 nm in the CD spectrum, while a positive peak at around 280 nm was observed in the CD spectrum of the unstructured ssDNA. Therefore, it is hypothesized that tandem G-hairpin and type 1 pre-GQ are generated in the interval-nhtel chains. Characterization by AFM showed loose

Figure 5 | Unfolding force curves obtained by AFM-based SMFS. (a) Schematic of setup for unfolding force measurement; (b) and (d) typical force-extension curves gained from unfolding; (c) and (e) histograms of unfolding forces; (b) and (c) nhtel in 100 mM KCl buffer; (d) and (e) interval-nhtel in 100 mM KCl buffer. At least 40 single-molecule stretching curves were used for the construction of each force histogram.

DOI: 10.31635/ccschem.021.202101436
CCS Chem. 2021, 3, 3192–3204
beads-on-a-string structures, further demonstrating successful folding of the interval structures (Supporting Information Figure S2). The single-molecule unfolding force measurements on the interval-htel were then carried out by following a protocol similar to that of the nthel. Interestingly, only force-extension curves characterized by a long-force plateau of $39 \pm 12$ pN (Figures 5d and 5e), which corresponds to the unfolding of partially folded GQs, were obtained.

This further supports the hypothesis that the tilted plateau with a mean force of $59 \pm 19$ pN gained in nthel is a result of the inter-G-quadruplex interactions, which further stabilize the folding structures. Indeed, the repetitive GQ sequences in the interval-htel ssDNA were separated by the ssDNA spacers, which inhibit the interactions between G-quadruplexes. In addition, the interactions between G-quadruplexes were confirmed by the long persistence length of the nthel chains ($6.8 \pm 1.0$ nm, means $\pm$ S.D.) compared to the interval conformation ($3.1 \pm 0.8$ nm, means $\pm$ S.D.) (Supporting Information Figure S11). This is highly consistent with the theory that rigid polymer chains have increased persistence length.43

The probability of obtaining a tilted plateau is 33%. Within that, the probability of the rupture events with higher force ($>55$ pN) is 44% (calculated based on the integral area of the Gaussian fitting curve, Figure 5c black). Therefore, in the long-telomeric sequences, the probability of QQI is at least 15%, which is higher than that observed in the short telomeric sequence (e.g., 144 nt) system ($\sim5\%$).43

According to the results obtained, both GQ and partially folded GQ were present in nthel under the experimental conditions, and the unfolding force of tandem partial GQs was approximately $40$ pN, which is higher than that of a G-hairpin but lower than that of a single GQ.42 Therefore, it is reasonable to deduce that G3 and G19 folded into tetraplexes, while G1 and G21 did not (Figure 4d). Additionally, the folding behaviors of G2 and G20 remained unknown. In this case, type 1 pre-GQ and type 2 pre-GQ were the dominant conformation structures of partially folded GQ in nthel. The CD spectra revealed that G-hairpin and type 1 pre-GQ are the dominant structures of long-chain telomeric DNA (nthel), while the SMFS results suggested that type 1 pre-GQ and type 2 pre-GQ are the dominant structures. For this reason, we concluded that type 1 pre-GQ with four vacant sites is the dominant conformation structure of long-chain telomeric G-chain in the KCl buffer.

The above conclusions/hypothesis were further supported by our DNAzyme activity assay. Recently, it has been reported that the catalytic properties of DNA G-quadruplexes rely on their structural integrity. The presence of a G-vacancy site not only decreases the stability of the G4 structure but also its hemin-binding affinity, which both agree in the overall lower catalytic performance.64 We investigated the catalytic activity of long-telomeric ssDNAs using the DNAzyme assay (Supporting Information Figure S12). The catalytic activity of the nthel and interval-htel telomeric ssDNA both have been improved after the addition of acyclovir (guanine derivatives). The DNAzyme activity assay demonstrates the existence of G-vacancy sites in the G-quadruplex formed by long-telomeric ssDNA, which further proves our hypothesis that there exist unfolded guanines at the 3’ or 5’ end of the long-telomeric ssDNA.54 In addition, the DNAzyme assay shows that in the absence of acyclovir the catalytic activity of nthel is much higher than that of interval-htel, which indicates that the structural integrity/stability of nthel is better than interval-htel. This finding is in good agreement with our single-molecule force spectroscopy results.

**Dynamic folding of tandem pre-G-quadruplexes**

To investigate the dynamic folding behavior of tandem pre-GQs, repeated stretching/relaxation was performed on one molecule using the AFM-based single-molecule force spectroscopy. To achieve that, RCA was used to synthesize long-telomeric DNA on the Au-coated substrate in-situ in a “graft from” fashion. An AFM tip was used to pick up a single molecule on the substrate and manipulate it repeatedly in stretching/relaxation cycles without rupture. Force-extension curves collected during the described cycles are shown in Figure 6 and Supporting Information Figure S13.

The stretching curves in a stretching/relaxation cycle exhibited force plateaus with a mean value of $29 \pm 7$ pN (blue, Figures 6a and 6c, and Supporting Information Figure S13a), which was approximately 10 pN lower than that of force plateaus gained during unfolding of nthel after complete folding (orange curve, Figure 6a and Figure 5c). The results demonstrated that the telomeric chains were unable to immediately refold into pre-GQ, resulting in a remarkable hysteresis between stretching and relaxation curves (the gap between the color and black curves in Figure 6a). Meanwhile, the plateau force of a stretched molecule did not increase even after relaxation for 1 min (Supporting Information Figure S14). The results indicated that telomeric DNA can fold into tandem G-hairpins after relaxation for seconds, whereas its folding into pre-GQ needs much longer time. In the single-time stretching experiment, long-chain nthel in the KCl buffer has approximately 10 min to fold before the first stretching curve is obtained (see details in “Experimental Methods” section). In the stretching-relaxation experiment, however, the tip cannot stay on the substrate for such a long time between two cycles. Hence, 40% PEG (w/w) was added into the buffer to accelerate this process. Typical force curves shown in Figure 6a (green curve) and S13B represent the characteristic folding/unfolding cycle of the long-telomeric ssDNA.
DNA in KCl buffer containing 40% PEG. Remarkably, the mean force obtained from the plateau region increased from \( \sim 29 \) pN in the absence of PEG to \( \sim 47 \) pN in the presence of PEG, suggesting that the high force plateau can reappear within seconds in the presence of PEG. Additionally, PEG is a mimic of the molecular crowding environment inside a living cell,\(^65,66\) and long-telomeric DNA can quickly (in seconds) fold in vivo to pre-GQ.

To exclude the QQI, interval-nhtel chains were manipulated repeatedly by the method described above. Surprisingly, the mean unfolding force decreased to 17 \( \pm 5 \) pN (blue, Figures 6b and 6d, and Supporting Information Figure S13c), indicating a loose structure compared to the structures formed by nhtel (Figures 6a and 6c) in a short time. And it was approximately 20 pN lower than that of force plateaus gained during unfolding of interval-nhtel after complete folding (orange, Figures 6b and 6d, and Figure 5e). After the addition of PEG, the mean force obtained from the plateau region increased to 38 \( \pm 13 \) pN (green, Figures 6b and 6d, and Supporting Information Figure S13d). These results suggested that the folding process of telomeric DNA was further decelerated in the presence of the interval sequences due to the absence of induction effects from neighboring telomeric sequences, and PEG can also accelerate the folding of interval-nhtel.

The thermodynamics of folding/unfolding of pre-GQ were investigated for a thorough understanding of the underlying mechanism. The stretching and relaxation curves were fitted by the WLC model, and the length change (\( \Delta L \)) caused by conformation transition was calculated. The variation in energy (\( \Delta E \)) during the folding/unfolding process was then estimated based on the area enclosed between the stretching and relaxation curves:

\[
W = \int_{A}^{B} F(x) \, dx
\]

\[
\Delta E = W_{\text{stretching}} - W_{\text{relaxation}}
\]

where \( A \) and \( B \) represent the start and end points of the force curve, respectively. \( W_{\text{stretching}} \) and \( W_{\text{relaxation}} \) were obtained from the stretching and relaxation curves, respectively. A plot of \( \Delta E \) as a function of the apparent \( \Delta L \) is shown in Figure 7a. It is clear that \( \Delta E \) increased with the increasing \( \Delta L \). Indeed, the slope of sample in the KCl buffer with PEG was 27 zJ/nm, which was nearly two times of the sample in the KCl solution with no PEG. This was consistent with the conclusion deduced by the SMFS study that pre-GQ was more stable in the presence of PEG, which can be attributed to the conformation transition of pre-GQ and GQ induced by molecular crowding. Additionally, the dependence of \( \Delta E \) on the loading rate has also been explored. As shown in Figure 7b, \( \Delta E \) increases with the increase of the loading rate, indicative of the nonequilibrium unfolding process of tandem G-quadruplexes.

As suggested by CD and SMFS results, the folding of tandem G-quadruplexes can be described as a two-step process involving partially and completely folded states. The long unstructured G-rich DNA (state I) was rapidly
assembled into the G-hairpin (state II) (Figure 8, step 1) with a mean unfolding force of 20–30 pN, which was measured in repeated stretching/relaxation cycles. After incubation for several minutes or addition of PEG, free guanines (red) folded into G-hairpin to allow generation of pre-GQ and complete GQ. The tandem pre-GQ exhibited a mean unfolding force of 40 ± 8 pN. In addition, inter-G-quadruplex interactions may be observed, resulting in further stabilized conformation (59 ± 19 pN). Transition from state II to state III (step II) requires at least 10 min.

Conclusion

We have reported a new type of G-quadruplex (pre-GQ) with multiple vacant sites in long human telomeric DNA and elucidated that this new type of G-quadruplex exhibited decreased stability compared to a complete G-quadruplex. The formation of pre-GQ may due to the influence of neighboring structures or physical restraints on long-telomeric ssDNA. According to the number of guanines involved in folding, pre-GQ can be divided into two types, which involves four (type 1) and two vacant sites (type 2), respectively. The combination of CD spectroscopy and SMFS results show that type 1 pre-GQ containing four vacant sites is a dominant structure. Furthermore, the results of CD spectroscopy and SMFS suggested a two-state folding process of long-telomeric ssDNA (nhtel), which involves conformation transition of three kinds of partially folded GQ (G-hairpin, pre-GQ type 1 and type 2) and complete GQ. The formation of G-hairpin from an unstructured telomeric ssDNA chain can happen in seconds, while the formation of pre-GQ takes around 10 min and could be accelerated by molecular crowding conditions. The present study deepens our understanding of the structure and dynamic folding/unfolding process of long-telomeric DNA. In addition, our findings may provide a new direction for the development of novel anticancer drugs, since such long-telomeric G-strand also exist in the double-stranded telomere DNA.25,46

Supporting Information

Supporting Information is available and includes all the DNA sequences (Tables S1–S3), additional AFM images,
CD spectra, DNAzyme activity assays, and SMFS data (Figures S1–S14).

Conflict of Interest
There is no conflict of interest to report.

Funding Information
This work was funded by the National Natural Science Foundation of China (nos. 21827805 and 21525418).

References


