THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 9499−9506

pubs.acs.org/JPCB

Folding Dynamics of Parallel and Antiparallel G-Triplexes under the Influence of Proximal DNA

Xi-Ming Lu,[†,‡] Hui Li,[*,†,‡] Jing You,[†,‡] Wei Li,[†,‡] Peng-Ye Wang,[†,‡] Ming Li,[†,‡] Shuo-Xing Dou,[*,†,‡] and Xu-Guang Xi[§,∥]

[†]Beijing National Laboratory for Condensed Matter Physics and CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China

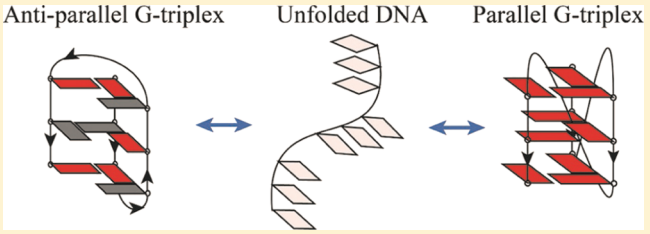
[‡]School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[§]College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, China

[∥]LBPA, IDA, ENS Cachan, CNRS, Université Paris-Saclay, Cachan F-94235, France

Supporting Information

ABSTRACT: The G-triplex is a kind of DNA structure formed by G-rich sequences. Previous studies have shown that it is an intermediate for the folding of G-quadruplex and has an antiparallel structure. The folding dynamics of this G-triplex structure, however, have not been well studied until now. In addition, whether a parallel G-triplex structure exists, remains unknown. In this study, by using single-molecule fluorescence resonance energy transfer and circular dichroism spectroscopy methods, we have studied the folding dynamics of the G-triplex and revealed at the single-molecule level, for the first time, that G-triplexes have both parallel and antiparallel structures. Moreover, we have investigated the effects of proximal DNA on G-triplex folding. We have found that both single-stranded TTA and double-stranded DNA at either end of a G-triplex sequence can reduce its folding speed. More interestingly, when located at the 5′ end of a G-triplex sequence, the proximal DNA will favor the folding of parallel over antiparallel G-triplex structures. As G-triplex is an intermediate for G-quadruplex folding, the present results may also shed new light on the folding properties of G-quadruplex structures, in terms of dynamics, stability, and the effects of proximal DNA.



INTRODUCTION

In human chromosomal telomere DNA, single-stranded TTAGGG tandem repeats at the 3′-overhangs are abundant.[1,2] This kind of single-stranded guanidine-rich sequence also exists in other chromosomal positions, such as the promoter regions of several proto-oncogenes.[3−5] Recent advances show that guanidine-rich sequences can form several kinds of structures such as G-quadruplex (G4),[6−8] G-triplex, and G-hairpin.[9−15] Among them, the G4 has been widely studied for its vital role in the chromosomal replication and as a drug-target candidate for cancer therapy.[16−19] G4 structures can be formed by a single oligonucleotide strand, stabilized by Hoogsteen-like hydrogen bonds and base stacking.[20,21] The G-triplex is generally recognized as an important intermediate in the multistep folding pathway of the G4. Recent single-molecule experiments using, for instance, magnetic tweezers and fluorescence resonance energy transfer (FRET) techniques, have revealed the roles of G-triplex in G4 folding.[12,22] Although the G-triplex structure has been reported in G4-folding experiments, few studies have been focused on its own properties. Besides the importance of G-triplex in G4 folding and telomere functions, G-triplex also exists in a thrombin binding aptamer (TBA) as shown by nuclear magnetic resonance (NMR) experiments.[23,24]

Previous studies have generally attributed the G-triplex conformation to an antiparallel type. For example, NMR studies resolved the antiparallel G-triplex in TBA structure, and interstrand G-triplex has been confirmed in solution state by atomic force microscopy.[11,14] Recently, circular dichroism (CD) spectroscopy experiments have reported that a parallel G-triplex structure may also form in positive ionic environments.[25] However, its existence is still undefined and lacking evidence at the single-molecule level. In addition, the folding dynamics of G-triplex have not been carefully investigated before. Especially, it is revealed that proximal DNA may influence the folding of G4,[26−30] but its effect on G-triplex folding is totally unknown.

In this report, by combining single-molecule FRET (smFRET) and CD spectroscopy methods, we have characterized the conformational dynamics of human telomeric G-triplex in KCl. We have demonstrated that both parallel and antiparallel G-triplex structures can form and undergo dynamic transitions. The G-hairpin as the intermediate of G-triplex folding is also revealed at the single-molecule level. Moreover,

Received: August 21, 2018
Revised: September 29, 2018
Published: September 30, 2018

DOI: 10.1021/acs.jpcb.8b08110
J. Phys. Chem. B 2018, 122, 9499−9506

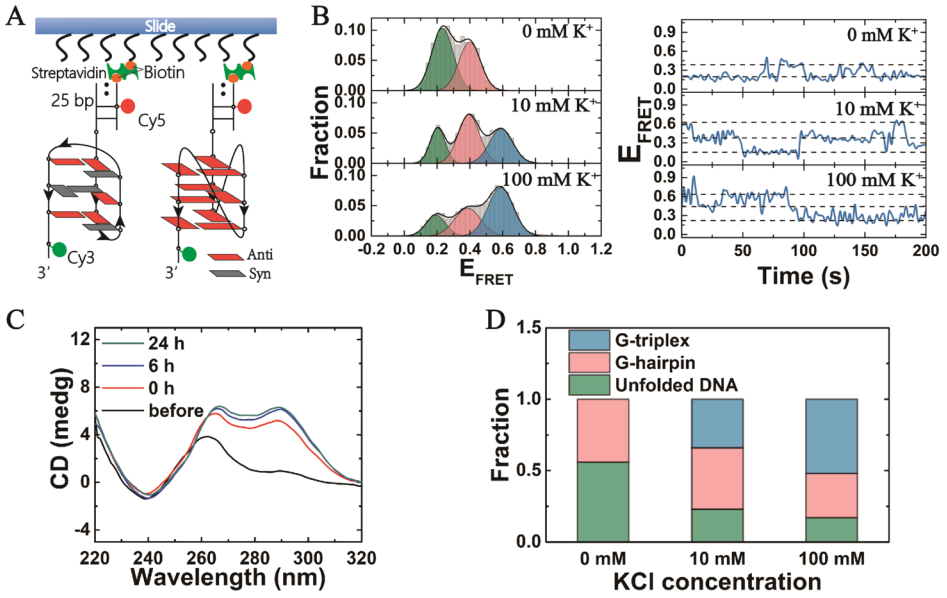


Figure 1. Observation of the conformational dynamics of G-triplex in KCl by smFRET. The donor fluorophore Cy3 was attached to the thymine of the second nucleotide from the 3′ end of the G-triplex sequence $(GGGTTA)_3$ and the acceptor fluorophore Cy5 at the third nucleotide from the 5′ end. The duplex stem was attached at the 5′ end of the G-triplex sequence. (A) Schematic representation of the two possible G-triplex structures (left, antiparallel; right, parallel). Anti and syn denote the different angles of the glycosidic bonds. (B) FRET histograms (left) and typical traces (right) in the imaging buffer containing different concentrations of KCl. Multipeak Gaussian distributions were used to fit the histograms. (C) CD spectra of a G-triplex sequence (GGGTTA GGGTTA GGG) measured before and at different times after adding 100 mM KCl. (D) Fractions of different folding structures at increasing concentrations of KCl, obtained from the areas of individual Gaussian peaks in (B).

we have found that the proximal DNA have great influence on the conformation and stability of the G-triplex. CD spectroscopy results showed that single-stranded TTA or double-stranded stem at either end of a G-triplex sequence will slow down its folding process. Furthermore, when located at the 5′ end, they will also reduce the proportion of the antiparallel conformation, which is further confirmed by smFRET experiments. The present results may deepen our understanding of the structure and folding of G-triplex and highlight the influence of proximal nucleotides. They may also shed new light on the folding properties of the G-quadruplex structures.

METHODS

DNA Substrate Preparation. All oligonucleotides required were purchased from Thermo Fisher. All DNA substrates with double-stranded stems were annealed by mixing single-stranded DNA (ssDNA) containing the G-triplex sequence and the complementary ssDNA at a molar ratio of 1:2 in a buffer with 20 mM Tris-HCl (pH 7.5). The concentration of ssDNA containing the G-triplex sequence was 10 μM. The annealing reaction was performed by incubating them at 95 °C for 5 min and then slowly cooling to room temperature for 6 h. All oligonucleotide sequences used here are listed in Table S1 in the Supporting Information.

Imaging Buffer for smFRET. A buffer containing 20 mM Tris-HCl (pH 7.5) with different concentrations of KCl was used. Glucose oxidase (1 mg/mL), 0.8% D-glucose, catalase (0.03 mg/mL), and 4 mM Trolox were mixed as an oxygen scavenger system to restrain photo-bleaching and photo-blinking. All reagents were from Sigma.

smFRET Data Acquisition. smFRET experiments were carried out using a self-built prism-type total internal reflection microscope. Cy3 was excited by a 532 nm laser (Verdi V2, Coherent). A water immersion 60× objective (NA 1.2) was used. Fluorescence signals from Cy3 and Cy5 were split by a dichroic mirror system and collected by an electron multiplying charge-coupled device (DU897, Andor). The coverslips and quartz slides were cleaned by a mixture of sulfuric acid, hydrogen peroxide, acetone, and sodium ethoxide. Then, the quartz slide surfaces were coated by a mixture of 99% mPEG (m-PEG-5000, Laysan Bio) and 1% biotin-polyethylene glycol (biotin-PEG-3500, Laysan Bio). Streptavidin (10 μg/mL) was added to the chamber and incubated for 8 min. After washing, 50 pM DNA substrate was added and incubated for 5 min. The unlinked DNA was then washed away by the imaging buffer. All smFRET videos were taken at an exposure time of 250 ms, for a total time of 200 s.

smFRET Data Analysis. The data analysis was carried out by scripts written in IDL and MATLAB, as was described earlier.[22,31] Briefly, pairs of Cy3 and Cy5 dots were identified from the experimental movies, and their intensity traces were measured. Then, the correlation coefficient between the intensity of the acceptor, I_A, and that of the donor, I_D, was calculated. The FRET efficiency, E_{FRET}, was calculated by $I_A/(I_A + I_D)$. Histograms of E_{FRET} were generated using over 100 individual molecules and fitted by multi-Gaussian distributions. The hidden Markov model was used to analyze the different states of E_{FRET}.

CD Spectroscopy. The CD spectra were recorded at room temperature with a Bio-Logic MOS450/AF-CD optical system (Bio-Logic Science Instruments, France) using a 1 mm path quartz cuvette with a reaction volume of 1.2 mL. The DNA concentration was 2.5 μM in the buffer with 20 mM Tris-HCl (pH 7.5). The CD spectra were measured before adding 100 mM KCl and at 0, 6, and 24 h after the addition of 100 mM KCl. The CD spectra were recorded in the UV regions (220–320 nm) with 1 nm increments. Each curve is an average of 10 repeated measurements.

DOI: 10.1021/acs.jpcb.8b08110
J. Phys. Chem. B 2018, 122, 9499–9506

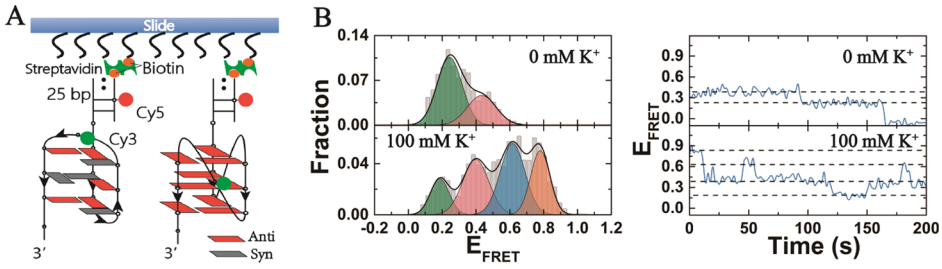


Figure 2. Observation of the conformational dynamics of G-triplex in KCl by smFRET, with Cy3 labeled at the middle of the 2nd loop. (A) Schematic representation of the two possible G-triplex structures (left, antiparallel; right, parallel) of $(GGGTTA)_3$. (B) FRET histograms (left) and typical traces (right) in the imaging buffer containing 0 or 100 mM KCl. Multipeak Gaussian distributions were used to fit the histograms.

RESULTS

G-Triplex Folding Process with an Intermediate G-Hairpin in KCl. To understand the folding process of G-triplex at the single-molecule level, we first characterized the conformational dynamics of G-triplex with human telomere repeats in KCl by smFRET. The substrate was constructed with an ssDNA sequence containing donor fluorophore Cy3 attached to the thymine at the second nucleotide from the 3′ end of the G-triplex motif. Its 5′ tail was hybridized with a complementary stem strand modified by biotin at the 3′ end for immobilization. In addition, the stem strand was attached by an acceptor fluorophore Cy5 at the third nucleotide from the 5′ end (Figure 1A). As the linker between the fluorophore and thymine is a single carbon chain, the fluorophores could rotate freely and orientate randomly. The measured FRET signal of the two fluorophores is the averaged result for the different orientations of dye dipole moments. Thus, it may report the distance change of the two fluorophores accompanying the conformational change of the G-triplex. The G-triplex DNA was surface-immobilized through a biotin−streptavidin bridge onto the polyethylene glycol-passivated microscope quartz slide. Then, the buffer solution containing 20 mM Tris-HCl (pH 7.5) with the indicated concentration of KCl and oxygen scavenger system was flowed into the chamber.

The FRET results of the G-triplex folding at different K^+ concentrations are shown in Figure 1B, together with the representative smFRET traces. We found that there were only two states around low FRET values of 0.2 and 0.4 without K^+. The third state with higher FRET values ($E_{FRET} \approx 0.6$) appeared with the addition of K^+, and its proportion increased with the K^+ concentration. According to the previous reports, cations such as K^+ are essential for G-triplex formation.[22,23,25] Thus, the above third state was expected to correspond to G-triplex folding in a K^+ environment. This was further confirmed by CD spectroscopy experiments where CD peaks at 260 and 290 nm were observed (Figure 1C and see below). As the peak with the lowest FRET value ($E_{0.2}$) was attributed to the unfolded state and that with the highest FRET value ($E_{0.6}$) to the folded G-triplex state, the peak with the medium FRET value ($E_{0.4}$) is very probably corresponding to a G-hairpin state, an intermediate for G-triplex folding. This is consistent with the previous reports.[11,14,22] The G-hairpin structure is suggested to form spontaneously as the first step in G4 folding.[32]

Fractions of the different structures were obtained by Gaussian fittings of the FRET histograms with three major peaks at 0.2, 0.4, and 0.6 (Figure 1D). The folded G-triplex is not stable enough in 10 mM K^+ environment (with a population of only 34%) but has a dominant population (52%) at 100 mM K^+ concentration. The proportion of the unfolded ssDNA and G-hairpin decrease gradually with the increasing K^+ concentration, but they still exist (17% for ssDNA and 31% for G-hairpin) at 100 mM K^+ concentration.

A previous study has shown that the Cy3/Cy5 dyes may stack on the end of the double-stranded RNA and stabilize the RNA structures.[33] To further check whether the dyes could influence the folding of G-triplex in our experiments, we elongated the G-triplex sequence at the 3′ end by adding three thymines and switched the location of Cy3 to the new thymine nucleotide that is farther away from the G-triplex structure (Table S1). The same results were observed (Figure S1), demonstrating that Cy3 labeled at the previous location (Figure 1A) has no influence on the G-triplex folding.

Both Parallel and Antiparallel G-Triplex Structures Were Observed. G-triplex was reported to be the intermediate for G4 folding previously.[10,12,22] In the folding pathway of G4, the ssDNA first forms the G-hairpin and the G-triplex consecutively before further folding into the final structure. Considering there are various G4 structures that may be formed in a K^+ environment, such as parallel and antiparallel types, it is expected that the G-triplex should also have both the parallel and the antiparallel types. However, the two folding structures of the G-triplex, if both indeed exist, cannot be distinguished by the smFRET measurements with the donor fluorophore Cy3 labeled at the 3′ end of the G-triplex sequence as shown in Figure 1A, because the distance between the donor and the acceptor remains, in theory, unchanged for the two structures.

Before trying to discriminate the two kinds of G-triplex folding at the single-molecule level by using a new DNA design, let us first go back to the measured CD spectra given earlier (Figure 1C). CD spectroscopy is widely used for obtaining the structural information of DNA substrates, by measuring the circularly polarized light absorption. The stacking of guanines in G-triplex can be thus characterized, which indirectly provides information on the strand orientations. It has been shown that parallel G4 with anti/anti guanines generates a positive CD peak at 260 nm, whereas antiparallel G4 with anti/syn and syn/anti guanines results in a negative peak at 260 nm and a positive one at 290 nm, in the K^+ environment.[34,35] In our CD experiments, we observed two positive CD peaks at around 260 and 290 nm, for the G-triplex sequence (GGGTTA GGGTTA GGG) in a 100 mM K^+ environment (Figure 1C). This indicates that the G-triplex structure formed is a hybrid of the parallel and the antiparallel types, because its CD spectra contain CD features of both the types. Moreover, when we examine the folding speed of G-

DOI: 10.1021/acs.jpcb.8b08110
J. Phys. Chem. B 2018, 122, 9499−9506

triplex in K^+, we may find that the CD spectrum right after adding K^+ is close to that at 6 and 24 h, which means that the ssDNA folds into G-triplex quickly, and the two types of structures also reach equilibrium quickly.

As the two types of G-triplex structures cannot be distinguished by the smFRET measurements with Cy3 labeled at the 3′ end of the G-triplex sequence, as just mentioned, we then designed a new DNA substrate by changing the position of Cy3 to the thymine of the 8th nucleotide from the 3′ end, so that the parallel and antiparallel G-triplex conformations would result in different distances between the donor and the acceptor (Figure 2A). The antiparallel type was expected to have a higher FRET value than the parallel one.

As shown in Figure 2B (0 mM K^+), there are two FRET states ($E_{0.2}$ and $E_{0.4}$) without K^+. As in the FRET experiment discussed earlier (Figure 1), the $E_{0.2}$ state represents the unfolded state of ssDNA. However, the $E_{0.4}$ state here should be different from the G-hairpin conformation observed in the earlier FRET experiment, as the new labeling position of Cy3 would result in a different FRET value if a G-hairpin is formed. As will be discussed below, the present $E_{0.4}$ state might represent an intermediate for the folding of the parallel G-triplex.

With the newly-labeled DNA substrate, the folding of the antiparallel G-triplex can result in a detected FRET signal with the highest value, because the donor is the nearest to the acceptor in this case (Figure 2A, left structure). As expected, in a 100 mM K^+ environment, a high FRET state at $E_{0.8}$ was observed (Figure 2B, 100 mM K^+), confirming the formation of an antiparallel G-triplex. Here, it should be pointed out that the $E_{0.8}$ state may also consist of a G-hairpin structure folded from the two telomeric repeats near the duplex stem. The reason is that this structure is very likely to be an intermediate for the antiparallel G-triplex and the distance between the two fluorophores is also the smallest. To confirm this, we performed a control experiment with a G-hairpin sequence labeled with Cy3 and Cy5 at its two ends. In this case, we indeed observed the same $E_{0.8}$ state (Figure S2).

After adding 100 mM K^+, in addition to the appearance of the $E_{0.8}$ state, another new FRET state $E_{0.6}$ also appeared. This implies that the parallel G-triplex was also formed at 100 mM K^+. The parallel structure (Figure 2A, right) would have a distance of the two fluorophores that is slightly larger than that of an antiparallel G-triplex (Figure 2A, left), thus corresponding to a slightly lower FRET state. This can be seen from a rough estimation of the distance of the two fluorophores from the molecule structures. The estimated distance for the parallel G-triplex is 2.9 nm, which is larger than that for the antiparallel G-triplex (1 nm) (Figure S3). From Gaussian fittings of the FRET histogram at 100 mM K^+ (Figure 2B), we obtained the proportions of the two high FRET states ($E_{0.6}$ and $E_{0.8}$) as 35 and 24%, respectively, indicating that the parallel-type structure is the major conformation of G-triplex in a 100 mM K^+ environment.

Effects of Proximal DNA on G-Triplex Folding Revealed by CD Spectra. As we know, in human telomeres, there exist many tandem repeats of TTAGGGG, so a G-triplex sequence is actually not an individual unit but has nucleotides at both ends.[1,2] To investigate whether the proximal DNA has influence on the G-triplex folding and structure, we first performed CD spectroscopy experiments for a G-triplex sequence, GGGTTA GGGTTA GGG, with additional nucleotides TTA at its 3′ or 5′ end. The results are shown in Figure 3. Comparing with the results for the G-triplex sequence without the additional TTA (Figure 1C), TTA at the

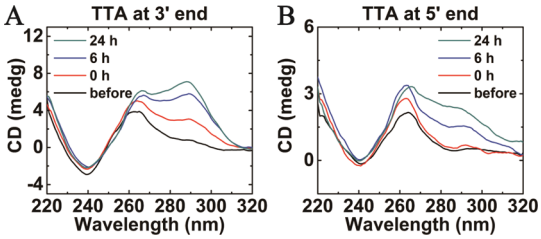


Figure 3. CD spectra of G-triplex sequence (GGGTTA GGGTTA GGG) attached with TTA, measured before and at different times after addition of 100 mM KCl: (A) TTA attached at the 3′ end; (B) TTA attached at the 5′ end.

3′ end of the G-triplex sequence obviously reduced the folding speed of the G-triplex, and the equilibrium was not reached until 24 h. The final structure of the G-triplex, however, remained almost the same, as can be judged from the spectra at 24 h in Figures 1C and 3A.

Interestingly, when TTA was added to the 5′ end of the G-triplex sequence, we found that the folding speed was reduced even more significantly and the positive peak at 290 nm was obviously much lower than that for the individual G-triplex without the additional TTA (Figure 1C). As the peak at 290 nm of the CD spectrum represents the antiparallel G-triplex structure, it suggests that the nucleotides TTA at the 5′ end can reduce the population of the antiparallel G-triplex. All these results show that the nucleotides TTA at either end of the G-triplex sequence may slow down the folding speed of G-triplex and, furthermore, TTA at the 5′ end can shift the equilibrium of structure distribution by partially inhibiting the formation of the antiparallel G-triplex.

Besides the nucleotides TTA, we note that, in the FRET experiments, a double-stranded stem DNA is generally added to one end of the G-triplex DNA. The duplex stem is necessary to avoid potential physical constraints to the DNA sequence from the slide surface. We wonder if the duplex stem also has some influence on the G-triplex folding. To do so, we examined the CD spectra of the DNA substrates constructed similarly as that used in the above FRET experiments (Figures 1A and 2A) but without any labeling (dye molecules and biotin); that is, a duplex stem was linked to one end of the G-triplex sequence $(GGGTTA)_3$. First, we examined the CD spectra of a DNA substrate with the duplex stem attached to the 3′ end of $(GGGTTA)_3$. In this case, the duplex stem only caused a decrease in the folding speed of G-triplex (Figure 4A),

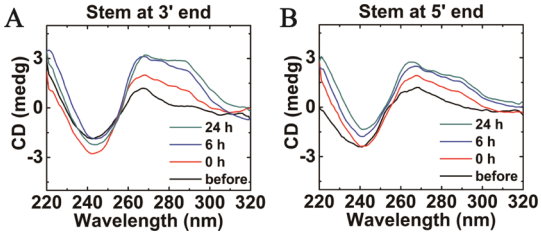


Figure 4. CD spectra of G-triplex sequence $(GGGTTA)_3$ attached with a double-stranded stem DNA, measured before and at different times after the addition of 100 mM KCl: (A) Stem DNA attached at the 3′ end; (B) Stem DNA attached at the 5′ end.

DOI: 10.1021/acs.jpcb.8b08110
J. Phys. Chem. B 2018, 122, 9499−9506

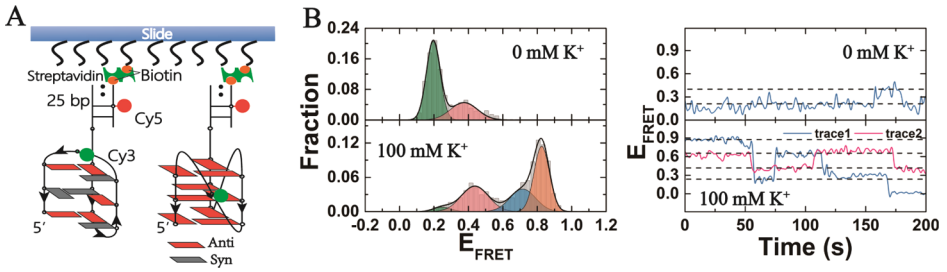


Figure 5. Observation of conformational dynamics of G-triplex at 0 or 100 mM KCl by smFRET, with Cy3 labeled at the middle of the 2nd loop and the duplex stem located at the 3′ end. (A) Schematic representation of the two possible G-triplex structures (left, antiparallel; right, parallel) of $(GGGTTA)_3$. (B) FRET histograms (left) and typical traces (right). Multipeak Gaussian distributions were used to fit the histograms.

which is similar to the case when TTA was at the 3′ end of the G-triplex sequence (Figure 3A). When the duplex stem was attached to the 5′ end of $(GGGTTA)_3$, not only the folding speed of G-triplex was decreased, but the fraction of antiparallel G-triplex (corresponding to the CD peak at 290 nm) was also reduced (Figure 4B), similar to the case when TTA was at the 5′ end of the G-triplex sequence (Figure 3B).

Taken together, the above results suggest that the duplex stem DNA and the single-stranded TTA have similar effects on the folding behaviors of G-triplex, and these effects depend on their positions relative to the G-triplex sequence. This finding should be important not only for designing DNA substrates in smFRET experiments, but also for studying the folding properties and functions of G-triplex in vivo.

Verification of the Effects of Proximal DNA on G-Triplex Folding by smFRET. The above ensemble-averaged results obtained by CD spectra revealed the significant influence of proximal DNA on G-triplex folding. To further confirm it at the single-molecule level, we performed the smFRET experiment using a DNA construct with the duplex stem at the 3′ end of $(GGGTTA)_3$ and the donor Cy3 on thymine of the 5th nucleotide (i.e., on the second TTA loop) from the 5′ end (Figure 5A). In this assay, the parallel and the antiparallel types of G-triplex can be identified, with similar FRET states to those obtained in the smFRET experiment described in Figure 2, where the duplex stem was at the 5′ end of $(GGGTTA)_3$. We found that the proportions of the two FRET states ($E_{0.2}$ and $E_{0.4}$) without K^+ (Figure 5B, 0 mM K^+) were similar to those observed earlier (Figure 2B, 0 mM K^+) (Figure 6). In the 100 mM K^+ environment, however, the proportion of the FRET state $E_{0.6}$ decreased from 35 to 25%, and that of the state $E_{0.8}$ increased from 24 to 43% because of the relocation of the duplex stem from the 5′ to the 3′ end of the G-triplex sequence. As the FRET state $E_{0.6}$ denotes the parallel G-triplex whereas $E_{0.8}$ consists of the antiparallel G-triplex and the intermediate G-hairpin, the proportion changes of the two FRET states indicate that the population of parallel G-triplex is reduced by the relocation of the duplex stem.

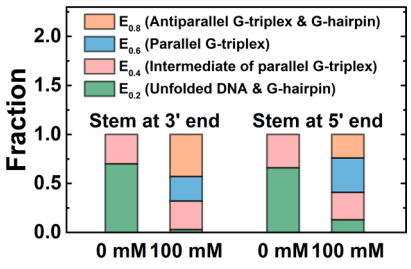


Figure 6. Fractions of different folding structures of the G-triplex sequence $(GGGTTA)_3$. Cy3 was labeled on the 2nd loop and the duplex stem was located at the 5′ or 3′ end. Fractions of different folding structures were obtained from the Gaussian fittings in Figures 2B and 5B.

In the $E_{0.8}$ state, we cannot determine the ratio between the antiparallel G-triplex and the G-hairpin. However, considering that this kind of G-hairpin is likely to be an intermediate for the folding of the antiparallel G-triplex, the increase in proportion of the $E_{0.8}$ state with the relocation of the duplex stem suggests that the antiparallel type is the preferred conformation of the folded G-triplex when the duplex stem is at the 3′ end of the G-triplex sequence. These smFRET observations are consistent with the previous CD spectrum measurements (Figure 4) showing that the fraction of antiparallel G-triplex (corresponding to the CD peak at 290 nm) is higher when the duplex stem is located at the 3′ end than at the 5′ end.

DISCUSSION

In this report, by using smFRET, we have studied the folding dynamics of the human telomeric G-triplex. By changing the labeling position of the fluorophore donor, we have revealed, at the single-molecule level, that G-triplexes have both parallel and antiparallel structures. Moreover, we investigated the effects of proximal DNA on G-triplex folding by using CD spectroscopy and found that both the nucleotides TTA and the double-stranded DNA located at either ends of the G-triplex sequence can decrease the folding speed of G-triplex. More interestingly, when the proximal DNA is added to the 5′ end, fractions of the two G-triplex conformations will change, with more parallel and less antiparallel structures. The results from the CD experiments were further confirmed by smFRET experiments.

Here, it should be noted that we have assumed that interstrand folding does not occur in all our smFRET and CD experiments. Although different G-strands may form tetramolecular or bimolecular G-quadruplexes at high concentrations,[36,37] a previous study has shown that the melting temperatures of the G-rich oligonucleotides including G-triplex sequence are independent of the DNA concentration (7−15 μM) in the presence of K^+, suggesting no intermolecular G-quadruplex or aggregation has formed.[25] In our experiments, the concentrations of the G-triplex DNA substrates in smFRET (50 pM) and CD spectroscopy (2.5 μM) are well below 7 μM.

Multistructures of G-Triplex and the Folding Pathways. On the basis of the smFRET and CD spectroscopy experimental results, we propose a mechanism for the folding

DOI: 10.1021/acs.jpcb.8b08110
J. Phys. Chem. B 2018, 122, 9499−9506

of the human telomeric G-triplex in a K^+ environment (Figure 7). For each of the two types of G-triplexes, there is a folding

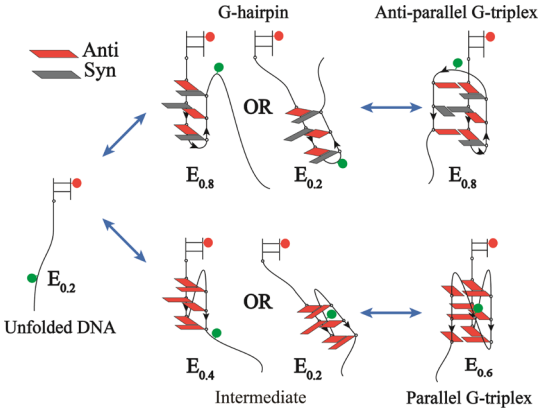


Figure 7. Proposed multipathways for the G-triplex folding of the human telomeric sequence $(GGGTTA)_3$. The upper pathway is for the folding of the antiparallel G-triplex structure and the lower one is for the folding of the parallel G-triplex structure. The FRET states of all the structures correspond to the case when the DNA substrate is labeled with Cy3 (green dot) on the 2nd loop and Cy5 (red dot) at the third nucleotide from the 5′ end, that is, on the double-stranded stem DNA which is attached at the 5′ end.

pathway. To form the antiparallel G-triplex (Figure 7, upper pathway), the unfolded ssDNA substrate (corresponding to the FRET state $E_{0.2}$) may first fold into G-hairpins, which have antiparallel structures.[11,15,22] For the DNA substrate with Cy3 labeled on the 2nd loop (as shown in the figure), the G-hairpin will manifest itself as FRET state $E_{0.2}$ or $E_{0.8}$, corresponding to the G-hairpin being away from the stem or close to the stem, respectively. Then, the G-hairpin further folds into the antiparallel G-triplex ($E_{0.8}$).

For the parallel G-triplex (Figure 7, lower pathway), there exists at least one intermediate structure in the G-triplex folding. The intermediate structure corresponds to the FRET state of $E_{0.4}$, which is between $E_{0.2}$ for the unfolded ssDNA and $E_{0.6}$ for the parallel G-triplex, as was observed in the FRET traces (Figure 2B). The structure of this intermediate is unknown; we suppose it is probably a parallel structure formed by the two GGGTTA tandem repeats, as shown in Figure 7, which corresponds to a distance of the fluorophore pair slightly larger than that in the parallel G-triplex structure ($E_{0.6}$). It is noted that the fraction of the intermediate state ($E_{0.4}$) decreases with the increase of that of the parallel G-triplex ($E_{0.6}$) with the addition of 100 mM K^+ (Figure 2B), which is further evidence of the existence of an intermediate in parallel G-triplex folding. Note that the existence of another intermediate structure for parallel G-triplex folding is equally possible. It is also a parallel structure of two tandem repeats but will manifest itself as FRET state $E_{0.2}$ because it is far away from the stem DNA and thus the two fluorophore molecules are more distant from each other. As this intermediate structure was mixed with the unfolded ssDNA in the FRET state $E_{0.2}$, it could not be observed separately in the smFRET experiments.

Now, let us reanalyze the experimental results in Figure 1 by referring to the proposed folding pathways in Figure 7. When Cy3 was labeled at the 3′ end of the G-triplex sequence, all the intermediate structures would correspond to the same distance between the two fluorophores, which is determined by a DNA duplex plus 6-nt ssDNA (i.e., GGGTTA). Thus, in the smFRET experiment, they would all belong to the same FRET state ($E_{0.4}$), as we have mentioned before. Similarly, the two G-triplex structures would correspond to the same distance between the two fluorophores and belong to the same FRET state. Note that this state has a FRET value of 0.6 as measured in the experiment (Figure 1B), but according to the G-triplex structures in Figure 7, this state should have a FRET value lower than 0.6 if Cy3 is labeled at the 3′ end. We think one possible reason for this discrepancy is that the TTA tail actually may not be totally free, but tends to bind in some way to the G-triplex structure, as it has been shown previously that the TTA tail can influence the folding of the G4 structures.[38]

Although antiparallel G-triplex structures have been proposed in previous studies using NMR, atomic force microscopy, FRET, and magnetic tweezers methods,[11,12,22,24] the parallel G-triplex structure is not confirmed yet at the single-molecule level. Our previous work also reported the G-triplex folding by smFRET,[22] however, the two topologies of G-triplex were not distinguishable due to the limitation of the fluorophore molecule positions. In the present study, we designed a new DNA substrate with Cy3 labeled on the 2nd loop in order to distinguish the two types of G-triplex structures. We confirmed, for the first time, the existence of parallel G-triplex at the single-molecule level. Coexistence of the two structures was also observed in our CD experiments.

As for the G-triplex folding pathway, our previous study only provided the folding pathway for the antiparallel G-triplex with the hairpin as the intermediate.[22] In the present study, we further defined another folding pathway for the parallel G-triplex, and observed an unknown parallel intermediate probably formed by two GGGTTA tandem repeats. More importantly, the multistructures and folding pathways of G-triplex may shed new light on the folding mechanisms of the G4 structures because G-triplex should be an intermediate structure in the folding of G4. In a K^+ environment, G4 have four different structures: antiparallel, parallel, hybrid I, and hybrid II.[27,39−41] The antiparallel G-triplex may be an intermediate for the folding of the antiparallel, the hybrid I, and the hybrid II G4 structures, whereas the parallel G-triplex may be an intermediate for the folding of the parallel G4 structure.

Effects of Proximal DNA on G-Triplex Folding. The influence of proximal DNA on the folding conformation and stability of G4 has been studied by using NMR, CD, UV-spectroscopy, and differential scanning calorimetry methods.[27−30,42] In our previous work, we found that less antiparallel and/or hybrid G4 was formed when the double-stranded stem DNA was at the 5′ end of G4, but the reason was not clear.[22] In the present study, for G-triplex, we have revealed that its conformational stability is also affected by the proximal DNA, including single-stranded TTA and double-stranded stem DNA. When the proximal DNA is located at the 5′ end of the G-triplex, it will partially inhibit the folding of the antiparallel G-triplex. Considering the intermediate role of the antiparallel G-triplex in the antiparallel and the hybrid G4 folding, the reduced antiparallel G-triplex could explain the less formation of the antiparallel and hybrid G4. Moreover, for the conformational dynamics, we further found that the proximal DNA located at either the 3′ or the 5′ ends of a G-triplex sequence can slow down its folding speed and the equilibrium state is not reached until 24 h (Figures 3 and 4). This is in contrast with the isolated G-triplex, which folds more quickly

DOI: 10.1021/acs.jpcb.8b08110
J. Phys. Chem. B 2018, 122, 9499−9506

on the addition of K^+ and the equilibrium is reached more rapidly (Figure 1C).

CONCLUSIONS

In summary, our smFRET and CD spectra experiments reveal more complex dynamic properties of G-triplex folding than ensemble experiments with untethered G-triplex sequences. The structural diversity of G-triplex and its dependence on the proximal DNA environments should be kept in mind when studying G-triplexes linked to single- or double-stranded DNA. In FRET experiments with G4, a double-stranded stem is usually attached to the G4 DNA substrate, which might be the reason for contradictions in the folding speed and conformation fractions of G4 between the smFRET experiments[43,44] and ensemble experiments using simply isolated DNA sequences.[45] It is worth mentioning that the human telomeric sequence has TTA at both the ends of each GGG sequence, thus, from our observations of the effects of TTA on G-triplex folding, there should be more parallel G-triplex and consequently more parallel G4 in vivo. This speculation needs future verifications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b08110.

Control experiment showing that the position of Cy3 dye has no influence on the G-triplex folding (Figure S1); conformation dynamics of the G-hairpin (GGGTTA GGGTTA) at 100 mM KCl (Figure S2); parallel and antiparallel G-triplex structures with duplex DNA used for estimating the distance between the two labeled fluorophore molecules (Figure S3); and sequences of DNA used in the experiments (Table S1) (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: huili@iphy.ac.cn. Phone: +86 10 8264 9682. Fax: +86 10 8264 0224 (H.L.).

*E-mail: sxdou@iphy.ac.cn. Phone: +86 10 8264 9484. Fax: +86 10 8264 0224 (S.-X.D.).

ORCID

Hui Li: 0000-0003-0551-9912

Ming Li: 0000-0002-5328-5826

Shuo-Xing Dou: 0000-0002-7201-2081

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project was supported by the National Natural Science Foundation of China (grant nos. 11774407, 11674383, 11474346), Key Research Program of Frontier Sciences, CAS (grant no. QYZDB-SSW-SLH045), and the National Key Research and Development Program (grant no. 2016YFA0301500).

REFERENCES

(1) Neidle, S.; Parkinson, G. N. The structure of telomeric DNA. *Curr. Opin. Struct. Biol.* **2003**, *13*, 275−283.

(2) Phan, A. T. Human telomeric G-quadruplex: Structures of DNA and RNA sequences. *FEBS J.* **2010**, *277*, 1107−1117.

(3) Phan, A. T.; Modi, Y. S.; Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Humanc-mycPromoter. *J. Am. Chem. Soc.* **2004**, *126*, 8710−8716.

(4) Le, H. T.; Miller, M. C.; Buscaglia, R.; Dean, W. L.; Holt, P. A.; Chaires, J. B.; Trent, J. O. Not all G-quadruplexes are created equally: An investigation of the structural polymorphism of the c-myc G-quadruplex-forming sequence and its interaction with the porphyrin tmpyp4. *Org. Biomol. Chem.* **2012**, *10*, 9393−9404.

(5) Trajkovski, M.; Morel, E.; Hamon, F.; Bombard, S.; Teulade-Fichou, M.-P.; Plavec, J. Interactions of pt-ttpy with G-quadruplexes originating from promoter region of the c-myc gene deciphered by nmr and gel electrophoresis analysis. *Chemistry* **2015**, *21*, 7798−7807.

(6) Bončina, M.; Lah, J.; Prislan, I.; Vesnaver, G. Energetic basis of human telomeric DNA folding into G-quadruplex structures. *J. Am. Chem. Soc.* **2012**, *134*, 9657−9663.

(7) Doluca, O.; Withers, J. M.; Filichev, V. V. Molecular engineering of guanine-rich sequences: Z-DNA, DNA triplexes, and G-quadruplexes. *Chem. Rev.* **2013**, *113*, 3044−3083.

(8) You, J.; Li, H.; Lu, X.-M.; Li, W.; Wang, P.-Y.; Dou, S.-X.; Xi, X.-G. Effects of monovalent cations on folding kinetics of G-quadruplexes. *Biosci. Rep.* **2017**, *37*, BSR20170771.

(9) Gray, R. D.; Buscaglia, R.; Chaires, J. B. Populated intermediates in the thermal unfolding of the human telomeric quadruplex. *J. Am. Chem. Soc.* **2012**, *134*, 16834−16844.

(10) Gray, R. D.; Trent, J. O.; Chaires, J. B. Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.* **2014**, *426*, 1629−1650.

(11) Rajendran, A.; Endo, M.; Hidaka, K.; Sugiyama, H. Direct and single-molecule visualization of the solution-state structures of G-hairpin and G-triplex intermediates. *Angew. Chem., Int. Ed.* **2014**, *53*, 4107−4112.

(12) Li, W.; Hou, X.-M.; Wang, P.-Y.; Xi, X.-G.; Li, M. Direct measurement of sequential folding pathway and energy landscape of human telomeric G-quadruplex structures. *J. Am. Chem. Soc.* **2013**, *135*, 6423−6426.

(13) Koirala, D.; Ghimire, C.; Bohrer, C.; Sannohe, Y.; Sugiyama, H.; Mao, H. Long-loop G-quadruplexes are misfolded population minorities with fast transition kinetics in human telomeric sequences. *J. Am. Chem. Soc.* **2013**, *135*, 2235−2241.

(14) Rajendran, A.; Endo, M.; Hidaka, K.; Teulade-Fichou, M.-P.; Mergny, J.-L.; Sugiyama, H. Small molecule binding to a G-hairpin and a G-triplex: A new insight into anticancer drug design targeting G-rich regions. *Chem. Commun.* **2015**, *51*, 9181−9184.

(15) Li, Y.; Liu, C.; Feng, X.; Xu, Y.; Liu, B.-F. Ultrafast microfluidic mixer for tracking the early folding kinetics of human telomere G-quadruplex. *Anal. Chem.* **2014**, *86*, 4333−4339.

(16) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-quadruplexes in gene promoters: A novel anticancer strategy? *Nat. Rev. Drug Discovery* **2011**, *10*, 261−275.

(17) Wells, R. D. Non-b DNA conformations, mutagenesis and disease. *Trends Biochem. Sci.* **2007**, *32*, 271−278.

(18) McLuckie, K. I. E.; Di Antonio, M.; Zecchini, H.; Xian, J.; Caldas, C.; Krippendorff, B.-F.; Tannahill, D.; Lowe, C.; Balasubramanian, S. G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *J. Am. Chem. Soc.* **2013**, *135*, 9640−9643.

(19) You, J.; Xu, Y.-N.; Li, H.; Lu, X.-M.; Li, W.; Wang, P.-Y.; Dou, S.-X.; Xi, X.-G. Helicase activity and substrate specificity of RecQ5β. *Chin. Phys. B* **2017**, *26*, 068701.

(20) Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988**, *334*, 364−366.

(21) Koirala, D.; Ghimire, C.; Bohrer, C.; Sannohe, Y.; Sugiyama, H.; Mao, H. Long-loop G-quadruplexes are misfolded population minorities with fast transition kinetics in human telomeric sequences. *J. Am. Chem. Soc.* **2013**, *135*, 2235−2241.

DOI: 10.1021/acs.jpcb.8b08110
J. Phys. Chem. B 2018, 122, 9499−9506

(22) Hou, X.-M.; Fu, Y.-B.; Wu, W.-Q.; Wang, L.; Teng, F.-Y.; Xie, P.; Wang, P.-Y.; Xi, X.-G. Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.* **2017**, *45*, 11401−11412.
(23) Cerofolini, L.; Amato, J.; Giachetti, A.; Limongelli, V.; Novellino, E.; Parrinello, M.; Fragai, M.; Randazzo, A.; Luchinat, C. G-triplex structure and formation propensity. *Nucleic Acids Res.* **2014**, *42*, 13393−13404.
(24) Limongelli, V.; De Tito, S.; Cerofolini, L.; Fragai, M.; Pagano, B.; Trotta, R.; Cosconati, S.; Marinelli, L.; Novellino, E.; Bertini, I.; Randazzo, A.; Luchinat, C.; Parrinello, M. The G-triplex DNA. *Angew. Chem., Int. Ed.* **2013**, *52*, 2269−2273.
(25) Jiang, H.-X.; Cui, Y.; Zhao, T.; Fu, H.-W.; Koirala, D.; Punnoose, J. A.; Kong, D.-M.; Mao, H. Divalent cations and molecular crowding buffers stabilize G-triplex at physiologically relevant temperatures. *Sci. Rep.* **2015**, *5*, 9255.
(26) Arora, A.; Nair, D. R.; Maiti, S. Effect of flanking bases on quadruplex stability and watson-crick duplex competition. *FEBS J.* **2009**, *276*, 3628−3640.
(27) Zhang, Z.; Dai, J.; Veliath, E.; Jones, R. A.; Yang, D. Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in k+ solution: Insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.* **2010**, *38*, 1009−1021.
(28) Mergny, J.-L.; Phan, A.-T.; Lacroix, L. Following G-quartet formation by uv-spectroscopy. *FEBS Lett.* **1998**, *435*, 74−78.
(29) Lim, K. W.; Khong, Z. J.; Phan, A. T. Thermal stability of DNA quadruplex-duplex hybrids. *Biochemistry* **2014**, *53*, 247−257.
(30) Krauss, I. R.; Ramaswamy, S.; Neidle, S.; Haider, S.; Parkinson, G. N. Structural insights into the quadruplex-duplex 3′ interface formed from a telomeric repeat: A potential molecular target. *J. Am. Chem. Soc.* **2016**, *138*, 1226−1233.
(31) Xi-Ming, L.; Hui, L.; Jing, Y.; Wei, L.; Peng-Ye, W.; Ming, L.; Xu-Guang, X.; Shuo-Xing, D. An optimization algorithm for single-molecule fluorescence resonance (smfret) data processing. *Acta Phys. Sin.* **2017**, *66*, 118701.
(32) Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H. Folding pathways of human telomeric type-1 and type-2 G-quadruplex structures. *J. Am. Chem. Soc.* **2010**, *132*, 14910−14918.
(33) Liu, Y.; Lilley, D. M. J. Crystal structures of cyanine fluorophores stacked onto the end of double-stranded RNA. *Biophys. J.* **2017**, *113*, 2336−2343.
(34) Karsisiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. Topological characterization of nucleic acid G-quadruplexes by uv absorption and circular dichroism. *Angew. Chem., Int. Ed.* **2011**, *50*, 10645−10648.
(35) Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. Circular dichroism and guanine quadruplexes. *Methods* **2012**, *57*, 64−75.
(36) Jaumot, J.; Eritja, R.; Tauler, R.; Gargallo, R. Resolution of a structural competition involving dimeric G-quadruplex and its c-rich complementary strand. *Nucleic Acids Res.* **2006**, *34*, 206−216.
(37) Vorlickova, M.; Chladkova, J.; Kejnovska, I.; Fialova, M.; Kypr, J. Guanine tetraplex topology of human telomere DNA is governed by the number of (ttaggg) repeats. *Nucleic Acids Res.* **2005**, *33*, 5851−5860.
(38) Tippana, R.; Xiao, W.; Myong, S. G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **2014**, *42*, 8106−8114.
(39) Dai, J.; Carver, M.; Yang, D. Polymorphism of human telomeric quadruplex structures. *Biochimie* **2008**, *90*, 1172−1183.
(40) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K + solution †. *Nucleic Acids Res.* **2007**, *35*, 6517−6525.
(41) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the Human Telomere in K +Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **2009**, *131*, 4301−4309.
(42) Ren, J.; Qu, X.; Trent, J. O.; Chaires, J. B. Tiny telomere DNA. *Nucleic Acids Res.* **2002**, *30*, 2307−2315.
(43) Aznauryan, M.; Søndergaard, S.; Noer, S. L.; Schiøtt, B.; Birkedal, V. A direct view of the complex multi-pathway folding of telomeric G-quadruplexes. *Nucleic Acids Res.* **2016**, *44*, 11024−11032.
(44) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938−18943.
(45) Marchand, A.; Gabelica, V. Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Res.* **2016**, *44*, 10999−11012.

DOI: 10.1021/acs.jpcb.8b08110
J. Phys. Chem. B 2018, 122, 9499−9506